POPULAR CONTROVERSIES IN WORLD HISTORY

Steven L. Danver, Editor

POPULAR CONTROVERSIES
in
WORLD HISTORY

STEVEN L. DANVER, EDITOR

Popular Controversies in World History

Volume One   **Prehistory and Early Civilizations**

Volume Two   **The Ancient World to the Early Middle Ages**

Volume Three   **The High Middle Ages to the Modern World**

Volume Four   **The Twentieth Century to the Present**

Popular Controversies in World History

INVESTIGATING HISTORY'S INTRIGUING QUESTIONS

Volume One
Prehistory and Early Civilizations

Volume Two
The Ancient World to the Early Middle Ages

Volume Three
The High Middle Ages to the Modern World

Volume Four
The Twentieth Century to the Present

Steven L. Danver, Editor

Popular Controversies in World History

INVESTIGATING HISTORY'S INTRIGUING QUESTIONS

Volume One
Prehistory and Early Civilizations

Steven L. Danver, Editor

ABC-CLIO

Santa Barbara, California • Denver, Colorado • Oxford, England

Contents

VOLUME THREE
The High Middle Ages to the Modern World

VOLUME FOUR
The Twentieth Century to the Present

Introduction

In the countless history courses I've taught over the years, the one question that invariably appears at some point (usually from a nonmajor who has to take the course for some general education requirement) is: "Why should we study history?" This is not an idle question, but an esoteric one that goes to the heart of what history is and what it can tell us. Usually, the student asking that question has the notion that history consists of a static set of "facts," unchanging (or, at least, it should not change) and ultimately meaningless for modern life. Students often buy into Henry Ford's famous take on the subject, "history is more or less bunk," rather than George Santayana's maxim, "those who cannot remember the past are condemned to repeat it." In the end, neither of these perspectives is especially true or helpful. This is because both writing history and understanding history are complex activities. They are our attempts to make sense of the past, usually drawing from incomplete or biased accounts of what actually happened. Even when the accounts are complete, the interpretations of history can vary radically depending on the perspective of the person writing. Perhaps the best explanation of the problem comes from the novelist Aldous Huxley, who, in his novel *The Devils of Loudun*, said "The charm of history and its enigmatic lesson consist in the fact that, from age to age, nothing changes and yet everything is completely different."

This work proceeds on the assumption that history is not a subject, but rather an activity. The activity of history engages the capability of students to use reason. On a purely anecdotal basis, I've asked many of my colleagues which skills they believed were the most important for their students to possess a high proficiency in when they begin college. Almost invariably, the two top answers were writing and critical thinking. In *Taxonomy of Learning*, developed in 1956 by Benjamin Bloom as an effort to show the evolution of mental skills in pyramidal form, critical thinking skills are integral to the third and fourth levels: application and analysis. The students who ask why it is important to study history are proceeding on the assumption that history is only an activity that engages the first two levels: knowledge and comprehension. If that were all there is to history, then Ford may have been right. However, application and

analysis are also key to understanding, without which one cannot reach the final two levels of Bloom's *Taxonomy*—synthesis and evaluation—which are essential to the creation of history. So, to summarize, critical thinking skills are key to moving from the first two levels—knowledge and comprehension—to the highest levels—synthesis and evaluation. In order to understand and, eventually, to write history, critical thinking is the transitional, indispensable skill.

Judging once again from my unscientific survey of my colleagues, it is one of the skills with which many students who are entering college struggle. This realization was the genesis of this project. *Popular Controversies in World History* takes as its subjects the topics over which there has been considerable historical debate. Some of these topics will not be familiar to students, but many of them will. Did the Great Flood, described in both the biblical book of Genesis and the Epic of Gilgamesh, actually happen? Is the lost continent of Atlantis just a myth, or was really such a place? Is the Shroud of Turin the actual burial cloth of Jesus Christ? Was William Shakespeare the sole author of all of the plays attributed to him? Who was the "man in the iron mask"? Did Franklin D. Roosevelt allow the Japanese attack on Pearl Harbor to happen as a pretext for the U.S. entrance into World War II? Did Lee Harvey Oswald act alone in assassinating John F. Kennedy in Dallas? These questions and more reveal that history is not a static set of facts, but rather a living, expanding set of ideas and interpretations. To understand those interpretations and formulate those ideas, critical thinking skills are paramount in importance.

The purpose of this work is to present the varying perspectives on events like these. These topics, as well as the ability to think critically about them, are vitally important parts of the social science curriculum at both the secondary and postsecondary levels. Each chapter takes a particular topic that has generated controversy either within the historical profession or in society as a whole and offers pro and con points of view, allowing readers to draw their own conclusions. The work covers all eras of human history, both before and after the advent of the written record. Each chapter in *Popular Controversies in World History* is formatted in the style of a historical debate, with a "pro" and a "con" section that presents contrasting perspectives. In most cases, both of these perspectives are or have been widely held within academia and supported by scholarship. The readers are then given the opportunity to exercise their critical thinking skills to evaluate the evidence presented by each side, to assess the validity of the arguments made by the authors, and eventually to determine which conclusions they accept or reject.

Of course, I could never have presented these arguments, ranging across so many eras and subdisciplines of history, by myself. This work represents the efforts of 62 other scholars with whom I have had the privilege to work. In addition, much of the early work on this project, especially determining the format to be used to accomplish our goals and formulation of the various questions

References

Bloom, Benjamin S., et al., eds. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: Longmans, Green, 1956.

Huxley, Aldous. *The Devils of Loudun*. New York: Harper, 1952, p. 259.

Santayana, George. *Life of Reason, Reason in Common Sense*. New York: Scribner's, 1905, p. 284.

Contributor List

Claire Brennan
James Cook University, Townsville, Australia

Justin Corfield
Geelong Grammar School, Geelong, Australia

Cheryl Golden
Newman University, Wichita, Kansas

Harald Haarmann
Institute of Archaeomythology, Luumäki, Finland

Peter N. Jones
Bauu Institute, Boulder, Colorado

László Kocsis
South East European University, Tetovo, Macedonia

John Lee
Utah Valley State University, Orem, Utah

Thaddeus Nelson
Columbia University, New York, New York

James Seelye
University of Toledo, Toledo, Ohio

Talaat Shehata
Columbia University, New York, New York

Olena Smyntyna
Odessa National University, Odessa, Ukraine

Barry Stiefel
Tulane University, New Orleans, Louisiana

Benjamin D. Thomas
University of Chicago, Chicago, Illinois

Patrick G. Zander
Georgia Institute of Technology, Atlanta, Georgia

I

Tool use is characteristic of hominids and apes, but not of other animal species.

PRO   Talaat Shehata
CON   Patrick G. Zander

PRO

If tool use is defined as being a "means to an end," then the contrary argument alluded to by the title of this chapter, that "evidence of tool use among other species tends to contradict this 'tool-making' as the characteristic of prehuman evolution," might also be entertained. But an investigation of the earliest evidence for the use of tools establishes, without any evidence to the contrary, that "only apes, hominids, and humans make true tools." To better understand and internalize this argument within a historical context, the visual senses needed to sift through the archaeological findings available will not merely suffice. Instead, by distinguishing what's been discovered in the material realm and how it was touched, felt, regularly used, and for what primary purpose by its original ape, hominid, and human creators will articulate their prehistoric drive for the use of the specific tools that, over time, they grew comfortable using. With that said, were there variances in the self-conscious uses of the tools by the apes, hominids, and humans (i.e., how did they differ or exhibit similarities)?

Great apes, with their variations, as evidenced by the existing chimpanzees, gorillas, orangutans, and bonobos, along with accumulated archaeological remains of the earliest hominids, provide profound insight into female and male *Homo sapiens* that evolved over millions of years. The earliest fossil records from around 1.9 million years ago identify the earliest of their ancestors, as the African *Homo erectus*. A common trait they all possess is the age-old practice of creating and using tools and passing these learned skills on to their offspring. Historical and fossil records also revealed that the late Miocene ape, *Oreopithecus bamboli*, predating most early hominids, discovered in Tuscany, Italy, had in the constructs of its hands firm pad precision gripping capabilities. This allowed for a significantly improved grasping ability. Through natural selection and evolution over tens of thousands of years, *Oreopithecus bamboli* was able to create and use tools that earlier apes had not been able to. Thumb and hand length and large, deep areas of insertion within the tendon of the flexor of the long thumb were not present in any earlier ape fossils. Precise gripping techniques have

I

Tool-using Chimpanzees

It has long been held that one of the distinguishing characteristics of humankind is the ability to create and use tools; however, according to a 2002 report in *Science* magazine, chimpanzees might have been using tools for much longer (Mercader et al., 2002). Archaeologists working in the Taï forest of Côte d'Ivoire found evidence that chimps transported stones to nut-cracking sites where they used them as hammers to process their foods. After placing a nut in a place where it could be immobilized, often in part of a tree stump, the chimpanzees then use granite stones to break them open. Additionally, a 2007 article in *Current Biology* reported that groups of chimps who learned to manipulate various devices in order to retrieve fruit were able to teach other chimps how to use the devices (Pruetz, 2007). Once a community had mastered a given device—for instance, a simple vending machine that would dispense fruit when a coin was deposited—and the device was transported to an unskilled community, the unskilled chimps were able to learn how to retrieve fruit simply by watching their skilled neighbors. This discovery suggests that chimpanzees made tools and may have passed that knowledge to new generations some 4,300 years ago, although scientists believe that the practice could have been occurring even earlier.

A chimpanzee uses a twig as a tool to catch termites. (DLILLC/Corbis)

been performed by great apes, cercopithecine monkeys, and nonhuman primates. This has included the often referenced "precision grip," which consisted of the use of a few finger tips, and most importantly, the thumb tip. Research and experimentation on nonhuman primates and humans has proven that what differentiated the humans from their near and distant primate cousins was that through evolution they acquired the unique ability to apply immense force to manipulate, use, and hold all different average-sized objects in a steady and secure manner between their fingers and the padded portion of their thumbs. This was a significant leap in evolution for humans, in their brain, hand, and tool-use coordination. The precision grip–acquired mode in fossil records, which dates back to over 3.5 million years ago, also shows that it was a feature that only hominids, as future humans, possessed. The hominids of that period are classified as the *Australopithecus afarensis*. There was no continual evolutionary bond between the *Oreopithecus bamboli* and the *Australopithecus afarensis*. This was proven by the discovery of fossils of the Miocene ape's hand bones. The epitome of hand-grasping precision had clearly developed over 5 million years earlier than the *Australopithecus afarensis* hominids. Gorillas and chimpanzees, on the other hand, despite the clear differences in their body mass and weight; the length of their hands in relation to each other, hominids, and humans; and their index finger and thumb ratios, have characteristically always been smaller, all sharing short thumbs. They did not then nor now possess the ability of the earlier Australopithecus and Oreopithecus hominids, along with later *Homo sapiens*' highly developed manipulative tool-using abilities and increased bipedal capabilities. These, over millions of years, had experienced selective evolutionary pressures that enhanced and favored those organs in each of these species, which in time (especially with humans) helped in the areas of hunting, farming, and harvesting. Therefore, regular evolving bipedality among hominids, and later humans, required the selective evolution of skilled hands. It is with the freeing of the hands and the strengthening of the hominids' and humans' lower calves, leg tendons, and muscles that these evolving and diverse species, would in time, be capable of becoming much more mobile in their daily affairs. In time, prior to the eventual arrival of modem humans, Australopithecus, Oreopithecus, and later more specifically, *Homo erectus*, became much more efficient in upright and sturdy walking. *Homo erectus* was using fire, manipulating and fine-honing its diverse collection of stone implements. Also, antler horns and other bone material were used for scraping and digging, with various boned appendages and objects used for engraving, sculpting, and stitching clothes. Gradually, they found imaginative and innovative uses for art forms and designs, as found in archaeological digs, such as ivory animal and human carvings, beads, small and large clay figurines, decorated and notched tools, and a diverse range of music devices. But, equally important, they left for further discovery in the next 20,000 years, as eminent scholars Louis and Mary Leakey had jubilantly discovered, striking, imposing, and brilliant cave paintings (Leakey, 1965). All that, with ever increasing

regular periods of physical activity invested for survival needs, moments of enhanced and often isolated contemplative thinking, social and cultural intra- and intergroup interactions, led, in time, to a higher individual and group consciousness, with a vivid sense of place, time, and community.

It was the Oldowan (referring to the prehistoric hominids, who had resided in the Tanzanian Gorge area of Olduvai during the Lower Paleolithic period) who used stone tools for gradually evolving industrial needs, as discovered by Louis Leakey. These tools of the period, which have been attributed to the Australopithecus, are regarded primarily as pebble creations, as a direct consequence of the pebble forms and shapes that they projected. The styles of these tools were very different from the tools of the Acheulean, which are regarded as their successors (Gonen, 2009). In the evolving quest to divide their labors, the Oldowans created scraping, grating, pounding, and chopping tools used for domestic uses. They were not the big-game hunting variety that one often equates with early and late Paleolithic hunting parties. These domestic tools were used for food finishing and processing needs. For over 2.5 million years these tools were meticulously used and have been and continue to be found by archaeologists, in most of the Olduvai Gorge area, where they had congregated and lived out their individual and collective lives. Approximately 1.5 million years later, and up to 500,000 years ago, Acheulan tools began to accumulate alongside the Oldowan finds, in some cases replacing them. They have often been unearthed by archaeologists in their numerous forays to the Olduvai Gorge area, alongside with many Oldowan tools. There is the possibility that most of these recent finds might establish much earlier dates for the existence and use of these tools by both the Oldowans and Acheulans. The gradual global reach of these tools, especially the Oldowan variety, has been found in Ethiopia, Kenya, South Africa, East Turkana, Iran, Pakistan, Israel, Spain, England, France, Germany, Italy, Cold War–era Czechoslovakia, Indonesia, Vietnam, China, and large parts of Asia.

So the argument that only apes, hominids, and humans made *true* tools, with the emphasis placed on the word true, can be clearly demonstrated in the construction of the tools. They fluctuated between rough and fine-grained two-sided cutting-edged instruments, often honed down to precise specifications for their specific use. Tools such as hand axes and various shaped chiseled stone spears were common instruments used by both the Oldowans and much later the Acheulans (Davidson and McGrew, 2006). Some scholars have even suggested that the level of both evolution and sophistication with their use of tools had progressed profoundly, despite the millions of years that separates them, to where a few of the more innovative had acquired a reasonable mastery of the art of raft building. This could advance clearer insight as to their migratory patterns and the regular unearthing of multivaried remnants of their tools and remains, within far-ranging global swathes of land. Of course, as of this writing, this theory has constructively continued to intrigue a few members of the scientific community, with little resolution in sight.

Most significantly, with the acquired necessity for true tool use over the past few million years and each historical event within its own context, modern and future humans would become the great beneficiaries. The evolution of the imaginary powers of the mind and intellect would begin to unleash forces and constructive and artistic beauty that the Oldowans, Acheulans, *Homo erectus*, and other animal species had sought and continue to seek. One could conclude that there is no way to equate ape's, hominid's, and human's true use of tools in nature's long historical evolutionary process. It is through this understanding of the natural evolution of art, instead of merely as cultural evolution, that we might be able to gain a stronger sense of how art could be studied and comprehended, possibly getting a better grasp of the long multitextured historical thread that ties the earliest apes and hominids to modern and postmodern humans.

As many curious and informed readers might already recognize, decades ago it was established that humans had chimpanzees as their closest relatives. They share 98–99 percent of identical genetic materials. Rats and mice have much fewer similarities to humans than humans have with chimpanzees. But, with minor differences in their genomes, the close cousins have in time diverged significantly in their evolutionary development. Coded within humans' genetic system are a few slight variances that determine one species' taste for thinking exercises such as philosophy, the arts, acquisition of language, mastery of agriculture, appreciation for music, and a quest for technological vastness. Chimpanzees, despite their close family and social ties, seem, through a quirk of biological nature, to have been left with much less "baggage" to contend with—whether that is a positive, negative, or neutral matter, that's left up to the reader to decide. It is within the amino acids located within the nuclei of the human cell that biological nature has arranged matters so that humans developed a much higher set of reflective and thinking skills that chimpanzees never acquired. Chimpanzees can't read, write, or speak, in the concentrated, focused, and persistent standards that humans have taken for granted. Chimpanzees are unable to research molecular biological matters, write symphonies, dramas, novels, and plays nor paint a profound multilayered art piece that might address the deepest cosmic, metaphysical, and cultural constructs and instincts that inform the group or social order they reside within (Sheeran and Shur, 2002).

For now, modern and postmodern men and women seem to have stolen the show. But for all the positives nature has bestowed upon them, with more complex linguistic, thinking, and social skills, nature has also endowed them with negative features, such as being more susceptible to diseases, such as Alzheimer's, viral hepatitis, malaria, and AIDS, all of which are diseases that do not afflict chimpanzees. It is also known that modern humans have a molecular form, through millions of years of evolution, that has been altered. That molecular form,

in the substance of sialic acid, has been found on the surface of all human cells. A single damaged gene in humans is the primary cause and code for that precise and profound change. Pathogens have used sialic acid as a docking system, which in that event, has allowed influenza and malaria strains to develop aggressively. It has been determined that chimpanzees and humans have between 10 and 60 genes separating them, which determine the state, substance, and content of the sialic acid in their biological systems. In each of these cases, humans lack these genes. That could possibly explain why chimpanzees are much less prone to acquire those disease strains than humans are. Being a random process, evolution has often, over the millions of years of ape, hominid, and human history, regularly reshuffled the biological and genetic decks whereby, with its continuous contact with the natural world, it allowed for a more "fit" species among its own and other species to materialize, possibly in time to be replaced by another more select and fit fellow or strange species.

In humans, *FOXP2*, a gene that has evolved over the past 200,000 years, has significantly helped in language and speech development. In comparison with other apes, the coded protein affected by *FOXP2* in humans was discovered through the regular sequencing of the amino acid in humans in relation to mice and apes, where the differences were significant in relation to that of chimpanzees (out of 715 locations it differed in only two). Despite being a miniscule difference, that difference allowed the development of speech and language in humans, but not in their closest cousins the chimpanzees. Also, a final and equally significant feature that differs between humans and chimpanzees was that of a mutant gene known as *MYH16*. This tiny gene mutation, which prevents the channeling or conveyance of a myosin variant, was found in what is known biologically and medically as chromosome 7. Chromosome 7's function impacts the production of myosin, a protein found in apes, chimpanzees, hominids, and humans, which aids in the contraction of muscle tissues. *MYH16* impacts the processes of biting and chewing, which would ultimately impact the functioning capacity of the jaw muscle of humans. Since humans lacked or possessed the slightest evidence of the mutant gene *MYH16* 2 million years ago, they were able to develop much smaller jaw muscles than other primates. A much disputed theory elaborated by Hansell H. Stedman of the University of Pennsylvania and a few fellow scientists states that humans, because of the lack of the *MYH16* mutant gene, were evolving with smaller jaw muscles and that the loss in muscle strength only increased the ability of their brains to grow much larger (Stedman, 2004). Critics contend that it was through natural selection and daily competition for food and the hunt, instead of more space provided for the brain to "increase" in, that humans were able to overwhelm their competitors, the apes and hominids. It seems clear that through a long, brutal, grueling, and fascinating process of evolution over millions of years, humans have found

themselves standing alone, for the time being, to assume the ever-so-fleeting evolutionary helm on this shared Earth (Sayers and Lovejoy, 2008).

With humans, based on their earliest tool use, their cultures have relied on a more methodical cumulative evolution. This gradual ratcheting process, which their ape and hominid competitors failed to exhibit, produced over an extensive period of history the modification of this cumulative evolutionary effect and further allowed for the creation, innovation, and handing down through successive generations greater and greater amounts of cultural artifacts. The evolution of material culture, with humans' growing grasp of how to create textile pieces and impute it with symbolic art, which would give meaning to their individual, tribal, and group aspirations, channeled their diverse and often limitless skills, energies, and emotions into a unique individual, tribal, and group representation of the world they had grown to spiritually and profoundly know. This would also find its unique artistic and symbolic expressions in wood and later metal items and through their diverse music and dancing rituals. Words, in much of human history, had not acquired the critical role they would play in the later earliest civilizations that arose in the Middle East, Asia, and other parts of the world. Later, historically, men and women would begin to develop a more complex and information-based system of binding their acquired material objects through the innovative and creative use of images and words. Humans, since the earliest of times, have always needed material objects to individually and collectively identify and uniquely differentiate themselves by. Artistically, within the human psyche, it was a means through which they could expand their mental, material, and spiritual horizons. Throughout time, humans have often felt oppressively bound to the limited physicality of their bodies, within the space and time that they had been born in. The central question seems to have been, "How might they accommodate themselves to their physical bodies and surrounding world and yet continue to maintain an untethered, dynamic, and engaging life?"

Unlike other animal species, humans' earliest tool use has evolved in a sophisticated manner over thousands of years, including private and public events where material objects were used to commemorate events that needed to be either forgotten or remembered. Also, death, loss, and absence are objectified through material objects. That explains why symbolic objects have taken on important and practical significance for weddings, births, secular and religious occasions, or when a child becomes an adult or a person dies. These are often very personal, public, and intimate moments for the participants. In other animal species, questionable tool use does not include fetish practices as seen with some humans, such as when a child compensates for the absence of a parent by clinging to a favorite soft stuffed toy, blanket, or object, which psychologically keeps the child close to the absent parent until he or she returns. Some family members may be fixated on jewelry or family heirlooms left to them by their mother, father, or other relative. These can be very intimate feelings, which other animal

species would never experience. The differences between species tool use and humans' evolutionary development over tens of thousands of years are too stark to be equated to one another, let alone to be ignored. They occur not only in any single geographic area, as with animal species, but in humans over tens of thousands of generations unfolded throughout the entire globe. After an extensive migratory pattern that took them to the farthest reaches of the world, these intelligent cultural, social forming, linguistically developed and developing beings, with their ever-advancing tool-using technology and abilities, excelled in very complex ways as to how to survive in their newly settled and resettled natural environments. Animal species, on the other hand, lived out most of their lives in a single geographic habitat, which they had overtime adapted themselves to. Once that habitat died or was destroyed, they perished with it. Unlike humans, all animal species' level of evolutionary adaptation was quite limited. Therefore, it is through the study of humans' material objects and artifacts that we are able to gain deeper insight into their thoughts and actions over history's long unfolding periods. As a conscious, self-reflecting, ever contemplative natural creation, animal species never had, nor ever will have, the deliberative, often profoundly intelligent thought processes and actions that humans have and continue to bring to their lives and the natural phenomena that regularly surround them.

Modern and postmodern man and woman have evolved their artistic predilections to include not only the art of objectifying and crafting artifacts but also infusing it with fragmented representations of each of their diverse groups, through poetic language, image and picture making, oral fiction and nonfiction narratives, humor-based oral renditions, and individual, family, and group histories. Their plate has always been full. Their art-making and expressive abilities would also evolve in the unfolding of scientific research, sports, religious and secular practices, and politics. In the ability to transcend one's worldly limitations, humans, through their diverse art forms, have been able to attain certain moments of rapture, which one doubts most animal species ever will attain. By being so artistically grounded, humans can regularly summon their art-oriented tool-using skills and abilities to attain yet other more compelling art pieces that summon their fellow self-conscious humans' readiness to be moved by greater, deeper, and more profound art work. The appreciation of beauty in its complex and multidimensional art forms has evolved into a human forte. That most abstract and practical of all notions that have been and are regularly created by more sophisticated human toolmaking and use is a reality that will never be experienced by any animal species in their constricted single and limited natural environments. Humans, through their tool-making and use over thousands of years, have attained a musical symphony conductor's virtuoso status among its fellow primates and diverse animal species. Through the highs and the lows of artistic evolution, humans have been able to challenge and interact with one another through their art. That is what has often provided them with the greatest

pleasure. This has and continues to be the ecstatic pleasure that their most ancient of human ancestors felt tens of thousands of years ago when they etched and painted their art forms on cave walls. This is also found in the pleasure that they obtained when they created their sculptures and figurines, beads, clothing, and tent and dwelling coverings. It is through the creation and appreciation of art, through the sophisticated manner of their tool-making and using skills and abilities, that humans were able to experience this most profound aspect of their existential presence in this world, that is, regularly answering the question posed about themselves, what is it that allows humans to gain regular and deeper insight into the window of their shared individual and collective humanity? No animal species will ever pose or ponder this question.

References and Further Reading

Aiken, Nancy E. *The Biological Origins of Art*. Westport, CT: Praeger, 1998.

Blocker, H. Gene. *The Aesthetics of Primitive Art*. Lanham, MD: University Press of America, 1994.

Currie, Gregory. *Arts and Minds*. New York: Oxford University Press, 2004.

Davidson, Iain, and William C. McGrew. "Stone Tools and the Uniqueness of Human Culture." *Journal of the Royal Anthropological Institute* 11, 4 (2006): 793–817.

Dennet, Daniel C. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster, 1995.

Gonen, Sharon. "Acheulian Giant-Core Technology: A Worldwide Perspective." *Current Anthropology* 50, 3 (2009): 335–67.

Karsten, Wentik. "Crafting Axes, Producing Meaning. Neolithic Axe Depositions in the Northern Netherlands. *Archaeological Dialogues* 15, 2 (2009): 151.

Latour, Bruno. "Visualization and Cognition: Thinking with Eyes and Hands." *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6 (1986): 1–40.

Leakey, Louis, et al. *Olduvai Gorge 1951–1961: A Preliminary Report on the Geology and Fauna*. New York: Cambridge University Press, 1965.

Mercader, Julio, Melissa Panger, and Christophe Boesch. "Excavation of a Chimpanzee Stone Tool Site in the African Rainforest" *Science* 296: 5572 (May 24, 2002), 1452.

Mullin, Molly H. "Mirrors and Windows: Sociocultural Studies of Human-Animal Relationships." *Annual Review of Anthropology* 28, 1 (1999): 201–24.

Pruetz, Jill D., and Paco Bertolani. "Savanna Chimpanzees, *Pan troglodytes verus*, Hunt with Tools." *Current Biology* 17:5 (March 6, 2007), 412–17.

Sayers, Ken, and C. Owen Lovejoy. "The Chimpanzee Has No Clothes: A Critical Examination of Pantroglodytes in Models of Human Evolution." *Current Anthropology* 49, 1 (2008): 87–114.

Sheeran, Lorri K., and Marc D. Shur. "Tree of Origin: What Primate Behavior Can Tell Us about Human Social Evolution." *American Journal of Human Biology* 14, 1 (2002): 82–84.

Sparshort, Francis. *The Theory of the Arts*. Princeton, NJ: Princeton University Press, 1982.

Stedman, Hansell H., et al. "Myosin gene mutation correlates with anatomical changes in the human lineage." *Nature* 428:6981 (March 15, 2004): 415–18.

VanDamme, Wilfried. *Beauty in Context: Towards an Anthropological Approach to Aesthetics*. Leiden, Netherlands: Brill, 1996.

Wilson, David Sloan. *Darwin's Cathedral: Evolution, Religion, and the Nature of Society*. Chicago: University of Chicago Press, 2002.

Wright, Robert. *The Moral Animal*. New York: Pantheon, 1994.

Wynn, Thomas, Frederick Coolidge, Steven L. Kuhn, and Mary C. Stiner. "Why Not Cognition?" *Current Anthropology* 49, 5 (2008): 895–97.

Zei'jong, Yu. "Archaeological Perspective on the Origins of "Behavioral Modernity." *Journal of Archaeological Research* 15, 1 (2007): 1–54.

CON

In 1965 Drs. Louis and Mary Leakey published their findings from years of research and excavation at Olduvai Gorge in Tanzania (Leakey, 1965). Among their findings were skeletal fragments of hominid species dating back 2.5 million years and a more recent genus that dated to just under 2 million years ago. The latter appeared to represent the crucial evolutionary branch leading to our own current species, *Homo sapiens sapiens*. The older remains came from the genus known as the Australopithecines, which remains the most ancient of generally acknowledged hominid branches. But, in addition to the remarkably old Australopithecine fragments, the Leakeys also discovered remains of that more recent hominid ancestor with decidedly more "human" facial and skull structure. This was *Homo habilis*, to date the oldest discovered ancestor of the *Homo* line, and its discovery contained another explosive correlation. Although the Australopithecine discoveries had contained no associated material finds, *Homos habilis* could be associated with the most ancient identifiable stone tools.

These earliest known forms of stone tools, apparently originating with *Homo habilis*, were termed "Olduwan" technology by the Leakeys after the Olduvai Gorge where they had been found. But, in the decades that followed, more examples would be found and associated with a successive species of hominid

that appeared to have descended directly from *Homo habilis*. *Homo erectus* (there are a few subspecies) emerged in Africa and eventually left that continent, radiating into the Middle East, Europe, and Asia. *Homo erectus* is the first species generally acknowledged to be a hunter of live prey and was remarkably durable, lasting from approximately 1.5 million years ago until only a few hundred thousand years ago in Asia. *Homo erectus* also is responsible for the first leap in technology, eventually improving techniques from the simple round choppers and hammers of Olduwan technology to the finely sculpted hand axes of the technical form known now as the Acheulean.

A hand-axe from the Acheulean culture from St. Acheul, France. (Goran Burenhult/StockphotoPro)

The archaeological record, then, would seem to confirm what is immediately visible in our own experiences—that humans use, manufacture, and improve tools while animals do not. With the accumulation of archaeological evidence of primitive stone technology, it would seem that human ancestors passed an all-important threshold, with the earliest iterations of the *Homo* genus gaining the ability to use and manufacture tools. It has, therefore, been speculated that tool use is the leading candidate, or one of the leading candidates, for what distinguishes our species as uniquely human. Further, the argument follows, while technology distinguishes and defines "man, the tool maker," tools also acted as a crucial trigger for natural selection and directed our evolutionary path. Expanded mental ability aided our ancestors in conceiving external objects as useful and helped them to modify objects to maximize their utility. Meanwhile, refined physiology, especially manual dexterity, allowed hominids to convert these thoughts into material reality. Finally, the use of specific tools themselves would emerge as its own selective pressure for evolution, with those born with better abilities to use butchering tools, for instance, better equipped for survival and the production of genetically similar offspring. This self-produced evolutionary pressure of tool-to-hominid relationship is sometimes referred to by experts as a "biofeedback loop."

This argument is a highly attractive one—one that appeals to our common sense and most of our experiences with the natural world and is supported by

the archaeological record. However, it would now appear that human evolutionary triggers are far more complex than this and that one of the crucial problems with this view is that tool use is, in fact, not at all unique to humans. Although tool use is clearly the exception rather than the rule among animal species, there are multiple examples of basic and even complex tool use in the animal world. Additionally tool use is not restricted to monkeys and apes; the use of tools for specific purposes is practiced by species in orders of animals with comparatively simple nervous systems, such as insects, crustaceans, and birds. By demonstrating the extent of tool use in the animal kingdom, it will become clear that tool use (and even manufacture) is not uniquely human, has not produced hominid-like evolutionary development in tool using animal species, and thus could not have been the crucial trigger for our distinctive development and definition.

Tool Use Definition and Study Methodology

Before examining the examples of tool use in other animal species, it is necessary to establish definitions for some of the most important terms. Defining *tool use* is notoriously difficult, especially as applied to use among animals. Some animal species use foreign objects in extremely simple ways that seem to test the boundaries of what is actually a tool. Some examples that are problematic would be an ape tossing its own feces as a defensive measure or smashing branches to fall on to potential predators. It follows that the term *tool* must itself first be defined. A tool is an external, or "foreign," object, not attached to the ground or "substrate," used for a specific purpose. There are disagreements about whether a tool can be internally produced by the user (such as in the case of the feces-throwing ape) and about whether an object could be animate or inanimate. Some monkeys, for instance, wield their own young to frighten away intruders, while some crustaceans brandish sea anemones to extend their own hunting ability. Benjamin B. Beck included both of these characteristics in the definition he published in his authoritative work on the subject, *Animal Tool Behavior*. For Beck, a tool could be internally manufactured or animate, but it had to be outside the user's body. *Tool use* then is the handling or manipulation of a tool (according to Beck's definition) in such a way as to accomplish a specific task. *Tool manufacture* is the production of tools for that specific task. It involves any modification of an existing material in order to adapt it to the purposes of the user.

There are some other important issues for consideration when examining tool use among animals and its documentation by humans. First, observing tool use among animal species in the wild is an extremely difficult and costly endeavor. Generally it requires systematic observation over extended periods of time in quite remote areas to ensure an undisturbed natural environment. This usually requires a structured study with staff, funding, and all that goes with such a project. While a tremendous amount has been done in this area, with the

enormous number of species in the world, there are a vast number that go unobserved under such conditions. Next there is the issue of corroboration and confirmation. There may be several first-time reports of tool use, but these sometimes cannot be confirmed by another example. Reports from rangers and hunters are somewhat common, for example, but these must be confirmed and documented in order to be considered a behavior genuinely associated with the specific species. Finally, there is the issue of behaviors observed among animals in captivity. These are understandably among the most common and systematic of behavioral observations, but there are questions about behaviors in such limited habitat, the different variety of objects for a species to examine, and also the issue of simple imitation. It is questionable then if observations in captive environments are truly relevant to the evolutionary process. Keeping these issues in mind, in the following list of examples of animal tool use, only confirmed and documented examples will be used of behavior observed in noncaptive habitats.

Examples of Tool Use among Animal Species

The most effective way to undermine the conception that tool use is uniquely human or is crucial to distinguishing hominids from other species is to list the most prominent and least-disputed examples. Tool use occurs across several orders, including insects, invertebrates, birds, and mammals. The variations of usage differ from the most basic to surprisingly complex, and some species even modify materials to create specialized or compound tools. There are also a multitude of different ways in which animals manipulate their tools. Beck identified 24 of these different "use modes." They include: *unaimed throwing* (such as tossing materials into the air, knocking off twigs to fall randomly on potential predators), *dragging, slapping, rolling, kicking, aimed throwing* (this generally involves throwing projectiles to injure or slow down prey, as well as defending against threats), *dropping* (releasing projectiles and using gravity to break or open targets, such as dropping rocks on to eggshells in order to eat the contents), *brandishing* or *waving* (such as waving away pests, in defense, or in courtship display), *prodding* or *jabbing, pounding* or *hammering* (most often to break open shelled food like nuts or shellfish), *inserting* and *probing* (generally searching for prey and other food sources), *prying* or *applying leverage, digging, balancing* and *climbing, propping* and *climbing, stacking, hanging* and *swinging, absorbing* or *sponging* (often to soak up nutritive liquids for practical transport and consumption), *wiping, draping* or *affixing, containing*, and *baiting* (such as placing food sources out to attract prey).

　　Examples of tool use have been reliably observed and documented since the 19th century, but serious work compiling multiple examples was not attended to until the work of Hugo van Lawick and Jane Goodall in the early 1970s. Their work, however, concentrated on primates and included only vertebrates. It was Beck who, in 1980, assimilated all the existing documented cases into a complete

catalog of animal tool behaviors. Most of the examples below will come from Beck's list and are attributed to him unless otherwise indicated. Some other sources, however, have made contributions, and there are some startling new developments, within the past year, concerning primate tool manufacture.

Invertebrates

Four classes have documented and confirmed cases of tool use: insects, crustaceans, cephalopods, and gastropods.

Neuropteran fly larvae: The neuropteran fly has a short lifespan and does not demonstrate tool use in its adult insect form. Neuropteran flies, however, spend up to two years as larvae (known as "ant-lions") and are proficient hunters at this stage, not relying on adults for food. Ant-lions dig funnel-shaped pits in the earth and then tunnel into the earth at the bottom of the pit to await prey that strays into the den. When prey arrives, the ant-lion seizes the prey and attempts to drag in underneath the soil to consume it. However, if the prey is able to escape the clutches of the ant-lion, as it attempts to scurry up the sides of the sandy burrow, the ant-lion grabs grains of sand and throws them rapidly at the escaping victim. These sand particles knock the escaping prey off the edge of the burrow and back to the bottom for the waiting ant-lion.

Dipteran fly larvae: The larval form of the dipteran fly is known as the "worm-lion," and although a quite distinct species from the neuropteran fly, it possesses virtually identical behavior in terms of tool use. The worm-lion also tosses sand projectiles into approaching or trapped prey, and like the ant-lion, it accomplishes this by rapid-fire "dorsiflexion" of its mandibles.

Myrmicine ants: Several varieties of ants (*Aphaenogaster rudis, A. treatae, A. tennesseensis, A. fulva*, and *Pogonomyrmex badius*) demonstrate a remarkable behavior in using natural material to absorb food for transport. After discovering a deposit of soft or liquid food sources, such as honey, fruit pulp, or the bodily fluids of prey, these ants will grasp pieces of leaf, bark, or wood and place them on to the food source. These "sponges" absorb the material, and the ants then carry the bloated leaf or wood scrap back to the colony, where it can be consumed by other worker ants. The ants consume the leaf or wood that is engorged with the sticky liquid and can easily ingest a much more nutritive meal. Although ants can carry pieces of leaf or wood quite proficiently, carrying liquid would be impossible without the device of using an absorbent for transportation.

Sphecine wasps: Sphecine wasps create burrows in the earth for their larval offspring and then catch and drop prey into the burrows for the larval wasp to consume. After provisioning the burrows for the final time before the larva is to evolve into its pupa form, the adult wasp uses earth to seal up the burrow. To do this, the adult wasp can use a variety of available tools to pound the soil until it is firmly packed. The wasp has been observed using bits of twig, bark, seeds,

and even remains of other insects to accomplish this task. The tools are held in the wasp's mandible and are tamped against the earthen wall, but at times the wasp may use a piece of twig to insert into the wall and probe. This may be to test the solidity of the packed wall as the process continues.

Hermit crabs: Hermit crabs are a well-known example of a crustacean that appropriates the empty shell of other deceased crustaceans. Hermit crabs do not enjoy the same protection of an exoskeleton that so commonly protects other crab varieties. Hermit crabs therefore seek out appropriate shells and affix themselves to the inside of the shell with the aid of a hooked leg or pleopod. Hermit crabs inhabit shells until they outgrow them and then search for another of more appropriate size.

Melia tessellata crab: This species of crab practices one of the most unusual, and admittedly borderline, uses of tools in the animal kingdom. The crab clutches anemones in its claws and keeps them there on a daily basis. When the crab is disturbed, it waves the anemones at the intruder to scare it away. But in addition to this brandishing behavior for display and intimidation, the crab uses the anemone to hunt as well. While hunting, the crab will allow the anemone to seize its prey item and then will reach over with its back appendages and appropriate the catch. When an anemone becomes fragmented or dies, the crab will discard it and attach another. This use of a living tool, manipulated with its fore legs, appears to be an adaptation, because its fore claws are not well developed.

Octopus: Octopi have been reported to use pebbles in the process of opening shells and removing the flesh inside. The stones are placed in the wedge where the shell opens in order to prop it open while the octopus consumes the animal occupant.

Fishes

Archer fishes: The only reported instance of fish using tools is one that may be questionable in terms of the definition one uses for a "tool." Toxotid fishes are well-documented shooters of water. The most commonly observed variety is the *Toxotes jaculatrix*, or the "archer" fish, of Southeast Asia. The archer fish takes in water through its gills and then shoots it out by compressing its gill covers. The water is channeled through a groove in the roof of the fish's mouth, while the fish presses its tongue against it to form a closed channel. The water shoots out under high pressure and can knock prey into the water. The archer fish is known to shoot water at prey resting on branches just above the surface and also at prey hovering above the water by flying.

Birds

Birds exhibit some of the most striking kinds of tool use and can be extremely efficient. One variety of bird also exhibits the behavior of modifying materials

> **Ravens' Reasoning**
>
> A study reported in the April 2007 *Scientific American* used tool use as a way to explore the cognitive abilities of ravens (Heinrich, 2007). Among other discoveries—that ravens can distinguish between and remember individuals (both raven and human)—researchers determined that ravens were able to reason out the cause and effect of simple tool use in advance. Older ravens were able to correctly complete a food–rewarded task on the first attempt, without the need for trial and error, by studying the arrangement of the task and determining the correct solution. Not only does it appear that tool use is not the exclusive province of primates, but other animals are not even limited to opportunistic tool use, and in some cases can reason about and remember utility and procedure.

to create a tool of appropriate size, which is an example of tool manufacture, although in its simplest form.

Egyptian vultures: The Egyptian vulture has been reported in van Lawick and Goodall's work as using stones to crack open ostrich eggs. There are two use modes reported for this. Egyptian vultures are known to select stones and carry them in their talons during flight, dropping them on to eggs below to crack the shells. They have also been documented in Christopher Baber's book *Cognition and Tool Use* as picking up stones with their beaks and flinging them at the eggs through quick neck flexion. Citing the work of other observers, Baber reports that Egyptian vultures earn approximately a 50 percent success rate in their rock-flinging efforts. A similar variety, the black-breasted buzzard of Australia, has also been reported as using the dropping technique, but this is circumstantial and unconfirmed evidence.

Woodpecker finches: The woodpecker finches of the Galapagos Islands display complex tool-using behavior indeed. Multiple reports have confirmed woodpecker finches securing cactus spines or bits of twig either by picking them off the ground with their beaks or by detaching them from the branches. The finches then manipulate the twigs with their beaks and insert them into openings in trees and bark to locate prey insects. Locating an insect, the finches have been known to lance and impale the prey, then remove it from the tree on the twig, or also to frighten the prey enough to force it out of the hole, where it can easily be plucked with the bird's beak. These finches have also been reported to modify the twigs by shortening them or by shaping the twig so that it will fit into small crevices, though this modification/manufacture behavior was displayed by a captive finch.

Mammals—Nonprimates

Elephants: Both African and Asian elephants have displayed multiple forms of tool use, generally using their trunks for grasping and manipulating. Observers

of African elephants in the wild have reported them to detach branches and limbs with their trunks and then, manipulating the limbs with their trunks, moving the sticks to irritated areas for scratching and to dislodge parasites like leeches. Others have reported seeing elephants in the wild propel projectiles like stones and logs with their trunks to repel and intimidate intruders. This has included the pulling up of fence posts and tossing them. Still others have documented Asian working elephants detaching and waving leafy branches to fan themselves and to repel unwanted flies.

Sea otters: Sea otters represent the best-documented case of nonprimate mammalian tool use, with several sources reporting of their tool usage in the wild, including Beck, Baber, and van Lawick and Goodall as well as Kathy Schick and Nicholas Toth in *Making Silent Stones Speak*. Sea otters, van Lawick and Goodall report, are the only nonprimate mammals to use stone tools habitually in their foraging and predation. Sea otters balance a stone on their abdomens while floating on their backs, grasp shelled mollusks in their paws, and strike the mollusk onto the stone until the shell fractures and they can conveniently consume the flesh. The stone is used as an anvil, but it may certainly be considered tool use because it is being carried and positioned by the user.

Bears: Several varieties of bears have been reported using tools such as limbs to knock down hanging fruit or pedestals for climbing. Most of these documented cases, however, are observations of bears in captivity. Baber, however, documents that polar bears especially will use rocks to drop or as projectiles to injure seals or walruses during attack.

Primate Mammals

Primates, including monkeys and apes, show a much wider variety of tool use and manufacture. In addition to more numerous examples among species, primates exhibit a wider variety of use modes and some clearly more sophisticated instances of manufacture. Although there are several examples of primate species exhibiting tool use, only the most prominent examples are listed below (Berthelet and Chavallion, 1993).

Howler and capuchin monkeys: Both howler monkeys and capuchin monkeys have been observed in the wild knocking off branches to fall below on to intruders, with some reports of this behavior also in spider monkeys and squirrel monkeys. This can be done by chewing off pieces, knocking them off with hands or tail, and shaking. Beck writes that capuchins were reported as having the most "apparent intent and instrumentality" in this task. In addition to issues of cognition, this may correspond with the capuchin's physiology, as they possess the greatest manual dexterity of the new world monkeys with greater prehensility and thumb opposability. Capuchins are also the only new world monkeys reported to jab intruders and competitors with a stick to prod them away and also to engage

in aimed throwing in their defensive behaviors. Captive capuchins have been observed positioning branches for climbing and reaching inaccessible areas and have also been documented as inserting sticks into tubes to probe and retrieve food items. Some experiments to test the tool-using potential of capuchins, however, suggested that capuchins did not have the cognitive ability or the dexterity to recognize and use stone technology.

Macaques: Macaque monkeys have been observed engaging in both more common and bizarre forms of tool use. Like several species of monkeys, macaques also routinely drop or throw objects downward to discourage unwelcome intruders. They are also reported to engage in aimed throwing of sticks and stones sometimes as part of courtship, sometimes as part of defensive or aggressive behaviors. The most controversial of their tool behaviors involves macaques approaching other males. One macaque is likely to pick up a baby and brandish it or wave it before another, which can have the result of diminishing aggression between the two adults.

Baboons: Baboons are among the most widely observed primates and display a number of tool-use behaviors. In the wild, baboons routinely use aimed throwing to fend off intruders. The throwing can involve stones, loose gravel, sand, and dirt and can be aimed specifically at the eyes of a threatening animal. Beck notes that this throwing is from a quadrapedal position or from a sitting position, as their physiology prohibits overhand throwing. Baboons have also been observed using long sticks to insert into wood to procure food or simply to extend their reach. Nearer to ape and early hominid tool use, wild chacma baboons use stones to smash open and hammer the tough-skinned or shelled fruits. They also use stones as a hunting weapon as when they smash scorpions to immobilize them before consuming them.

Orangutans: Orangutans also routinely break off and throw down branches to discourage and frighten away intruders. Although Beck comments that their aim is poor, this behavior still qualifies as aimed throwing. There have also been observations of wild and semiwild orangutans using sticks as a weapon to strike another orangutan or to kill a snake. Wild orangutans have also been observed to use leaves as a wiping tool to sponge away feces or saliva and even to cover ant hills to avoid bites. Interestingly, captive orangutans display far more tool use and ingenuity in captive situations, which possibly extends from their determination to escape their captive habitat (such as creatively breaking out of the London Zoo by crashing a potted plant through a skylight). Orangutans have remarkable receptivity to training and intuition, as well as a keen ability to imitate.

Gorillas: Gorillas routinely use unaimed throwing to frighten and intimidate intruders such as human observers, which includes sticks, branches, and leaves. They are also known to wave and brandish large sticks or poles in these acts of display. Gorillas are known to commonly engage in aimed throwing in captivity

that can include materials such as vegetables, water, and feces. Such powerful animals throw tremendously hard and project these items with remarkable and dangerous velocity and accuracy. Gorillas have been reported, but not in confirmed or published reports, using sticks as extenders to reach inaccessible food. This behavior has been recorded in adults in captivity.

Chimpanzees: The animal displaying, by far, the most examples of tool use in the widest variety of use modes is the chimpanzee. Chimpanzees have also enjoyed some of the most intense and outstanding scientific observation in the wild such as in the famous work of Jane Goodall in Gombe National Park (Goodall, 1986). Goodall has recorded a significant variety of tool uses and modes among chimps. First, like gorillas, chimps use tools such as branches and sticks in display behavior, mostly to intimidate. When display does not work and chimps fight among themselves, they use a variety of physical techniques like hitting and biting, but tools can also be used as weapons. Up until recently, this observed tool-use behavior has only included aimed throwing of stones, loose gravel, sticks, or bits of fruit and vegetation. This could happen in fights between chimpanzees or if chimps attempted to frighten away potential predators. Chimpanzees have also been observed in the wild quite dexterously using long, thin pieces of bark or stiff blades of grass to probe into holes in a log or into termite or ant dens to remove the insects for food. They insert the long leaf or bark into the den and allow the insects to crawl on it. Once they do, the chimp removes the leaf and consumes them. Chimps also, as first documented by Goodall, use stones as both hammers and anvils to open fruits and nuts and use sticks as hammers as well. Chimpanzees will break off pieces until they arrive at one of the appropriate size, if possible, which is a basic form of tool manufacture. But new studies are opening exciting windows into the tool manufacturing abilities and innovative tool uses of chimpanzees.

A recent study by Jill Pruetz and Paco Bertolani in the journal *Current Biology* documents remarkable new dimensions of chimpanzee tool use in hunting in Senegal. They have observed chimps detaching living branches and fashioning them into pointed spears, generally with their teeth. The pointed spears are then used to stab small game as it hides in crevices in trees or dead logs. Chimps using these spears jab them into the openings forcefully and withdraw the spear, tasting for blood or in some instances withdrawing small bushbabies impaled on the spear. The manufacture of these spears, according to Pruetz and Bertolani, appears to be a skill perfected and passed on to the young by mothers rather than male chimps. This is a newly documented behavior, never confirmed before, and the possibility exists that humans have been able to witness a newly developed skill and stage in the evolutionary process. Additionally Pruetz and Bertolani see possible implications for the importance of females in the development, transmission, and evolution of early technology.

Conclusion

The list of tool-use behaviors above should make clear that there are a great number of examples of usage in the animal world across a wide spectrum of animal orders. From our closest animal relatives, the chimpanzees, to animals as simple as a fly, tools are routinely used by these species in myriad ways. Explaining the appearance of these tool uses is the next task and is far more difficult. With primates, authors like Baber have been able to determine, for instance, the extent of cognition in chimps associated with tool use and their ability to recognize materials as possible tools to be adapted to specific tasks. In the orders of creatures with very simple nervous systems, however, this will remain a difficult and elusive pursuit.

One important corollary to the expansion of brain size and cognition among primates and hominids is the development of the utility and sensitivity of the hand. In her article on hand structure and tool development, "Primate Hands and the Human Hand," Francoise K. Jouffroy suggests not only that increased dexterity among primates contributes to the ability to manipulate tools in more diverse ways, but also that the shape and function of hominid hands would have provided the models, themselves, for tool development (Jouffroy, 1993). Observing the functions of the hand cupping to carry water, for example, would have provided the earliest inspiration for primitive bowls, according to Jouffroy.

It is almost always true that the closer the animal species is to humans, the more complex the tool behavior it exhibits. Although this is true, it is also true that tool use exists at many different levels in species other than humans. That being the case, it cannot be a tenable assertion to suggest that tool use or tool making are exclusively the province of hominids, that these skills and tendencies are the chief differentiator of our species, or that tool use was the crucial evolutionary trigger that sent hominids on the path to our own day. Tool use, in fact, cannot even be seen as a crucial evolutionary determinant in the animal kingdom. It is true that there are several species that demonstrate tool use, but there are far more that do not, many of which inhabit the same natural environments as those that have adapted tool use. Why haven't tool-using animal species come to dominate their respective environments? Although tool use and manufacture are clearly quite important in hominid development, they cannot be treated as "definitively" human, and for those that continue to seek such defining characteristics, they will likely have to consider broad combinations of material skills, physiological attributes, and cognition.

References and Further Reading

Baber, Christopher. *Cognition and Tool Use: Forms of Engagement in Human and Animal Use of Tools*. London: Taylor and Francis, 2003.

Beck, Benjamin, B. *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*. New York: Garland STPM Press, 1980.

Berthelet, Arlette, and Jean Chavallion, eds. *The Use of Tools by Human and Non-Human Primates*. New York: Oxford University Press, 1993.

Goodall, Jane. *The Chimpanzees of Gombe*. Boston: Houghton Mifflin, 1986.

Heinrich, Bernd and Thomas Bugnyar. "Just How Smart Are Ravens?" *Scientific American* 296 (April 2007), 64–71.

Jouffroy, Francoise K. "Primate Hands and the Human Hand: The Tool of Tools," in *The Use of Tools by Human and Non-Human Primates*, Arlette Berthelet and Jean Chavaillon, eds., Oxford: Clarendon Press, 1993, pp. 6–33.

Leakey, Louis, et al. *Olduvai Gorge 1951–1961: A Preliminary Report on the Geology and Fauna*. New York: Cambridge University Press, 1965.

Pruetz, Jill, D., and Paco Bertolani. "Savanna Chimpanzees, *Pan Troglodytes Verus*, Hunts with Tools." *Current Biology* 17 (2007): 412–17.

Schick, Kathy, D., and Nicholas Toth. *Making Silent Stones Speak: Human Evolution and the Dawn of Technology*. New York: Simon and Schuster, 1993.

Van Lawick, Hugo, and Jane Goodall. "Tool-Using in Primates and Other Vertebrates." *Advances in the Study of Behavior* 3 (1970): 195–249.

2

Agriculture, or the domestication of plants, diffused from its start in the Middle East to the rest of the world.

PRO   Olena Smyntyna
CON   Harald Haarmann

PRO

The origin and spread of agriculture over the world traditionally remains one of the most alluring and unequivocal questions in contemporary science. Searches for logical and reliable answers for questions about spatial and chronological frames of land cultivation practices and invention and about the time and place of a certain plant's domestication constitute an essential part of a series of hot contemporary discussions, among which one can mention topics of global historical context (such as the *Ex Oriente lux* concept, the problem of social inequality origin, the role of migrations in early human history, the reliability of the Bible as a source of historical information) alongside strictly subjective case studies (i.e., Great Flood hypothesis, the Fertile Crescent and cradle of civilization concept, gene flow theory, and the origin of Indo-Europeans).

The contemporary discussions about the origins of agriculture are also driven by their broad interdisciplinary context, which inevitably involves specialists in the large spectrum of social and natural sciences, such as archaeology; history, cultural studies, biology, and medical anthropology; human genetics; paleobotany and plant selection; geography; marine geology; demography; and many others. Most of these disciplines have developed independently during the past decades, with their theoretical achievements and new empirical databases open to novel perspectives for fresh insight into the problem.

This section will examine the migratory, or diffusive, approach to the problem of agriculture origin and plant domestication. In order to clarify its reliability and cognitive potential, we first must provide an overview of the history of the migratory version of the spread of agriculture, sum up theories and concepts proposed by representatives of different sciences, and synthesize the arguments put forward since W. G. Childe initiated the discussion of this idea. Following this historiographic approach, it will place the discussion in the context by examining recent trends, and special attention will be given to the problem of human migration versus the idea's diffusion, which caused the dissemination of the agricultural

23

mode of life worldwide. Finally, a synthesis concerning the connection of agricultural diffusion with the dispersal of Indo-European peoples will be made and corrolated with the Black and Mediterranean seas' level change at the beginning of Holocene period.

Migration and Diffusion: An Attempt of Identification

Migration could be defined as the total or partial change of location (habitat) or movement into new areas for certain periods of time or permanently. The term "migration" is widely applied in the social sciences as well as in the frames of biology, geophysics, astronomy, and computer sciences with reference to plants and animals, fish and birds, insects and cells, planets, systems, and data.

In contemporary social sciences, migration is interpreted as population displacement (translocation) and is usually viewed as one of the four basic genres of human activity alongside habitation, storage, and creation. Migration is an integral part of the professional terminology of contemporary archaeologists, ethnologists, sociologists, demographers, cultural anthropologists, geographers, and representatives of other social sciences. The sphere of its application and meaning seems to be so clear that some reference encyclopedias and dictionaries consider it unnecessary to define or interpret this concept. Nevertheless, more careful analysis of the concept of migration's application or sphere testifies to the fact that it is often applied to the process, which principally differs in spatial and chronological scale as well as in ecological, livelihood, ethnic, and social consequences.

The migration concept is applied in studies of the displacement of population (group and individual movements) as well as to the dispersion of the artifacts created by the people and their culture in general. Sometimes anthropological and other data indicate that culture transmission does not accompany its human substrate displacement; in such a case migration of ideas is assumed.

The possibility of culture migration was widely discussed in cultural anthropology at the beginning of the 20th century in frames of a wide variety and diffusion of schools and theories. The notion of cultural diffusion as a spatial transference of cultural phenomena was put forward by anthropologists, and this context of human history was interpreted as a series of cultural clashes, adoptions, and transfers. Long-distance contacts such as international trade and exchange, conquests, and conscious imitation were regarded as basic ways by which certain cultural phenomena or artifacts could surmount impressive distances from the point of their primary origin.

Origin and rapid upsurge of genetics during the second half of the 20th century provided a possibility to verify the hypothesis about ideas and artifacts translocation without human displacement on the basis of comparison of gene structure of population of certain areas. Such DNA spatial distribution studies, initiated at the turn of the 21st century, were brilliantly developed by Luigi Cavalli-Sforza

---

Hunter-Gatherer Food Storage in Jordan

Sometimes technology has to be in place before innovation can proceed. Such appears to be the case for the Stone Age agricultural revolution in the Middle East. Archaeologists in Jordan have recently uncovered a very large granary for the storage of the wild grains that were among the staples of the hunter-gatherer diet. Dating to approximately 9,300 BCE—about 1,000 years before the first signs of domesticated, settled agriculture—the facility unearthed near the Dead Sea town of Dhra' provides evidence that hunter-gatherers were able to establish settled towns, complete with communal food storage and work patterns, even before the development of domesticated agriculture. Researchers have speculated that these wild grains—including barley, lentil, and oats—were the progenitors of the domesticated plants that would flourish during the agricultural revolution.

Source: Bower, Bruce. "Ancient Granaries Preceded the Agricultural Revolution." *Science News* 176, 2 (July 18, 2009), 13.

---

and have resulted in a series of gene flow theories that have brought studies of human migration into historical retrospection on a principally different level.

Oasis Theory of W. G. Childe

Searches for logical and reliable explanation of agriculture origin can be traced back as early as the beginning of 20th century, when lengthy discussions about the historical and cultural background of the transition from a hunter-gathering economy to land cultivation and cattle breeding had been started in the course of intensive archaeological explorations and excavations in the Middle East. From the very outset the stumbling stones have been the definition of time and place of productive economy origin (with permanent rivalry of monocentric versus polycentric understanding of this cultural and economic phenomenon) along with the reconstruction of the mechanisms, pace, and ways of spread of agriculture in the world, as well as the chronology of these processes.

One of the earliest versions of the migratory explanation of agriculture origin and dissemination was proposed by William Gordon Childe in connection with his concept of Neolithic revolution, as early as the mid-1920s. According to Childe, drought and supply shortage stimulated food production in certain oasis areas characterized by a peculiar set of geographic, environmental, and social parameters that enabled (or even required) transition to new forms of livelihood and subsistence strategy.

His localization of this oasis in the Middle East at the beginning of the 20th century matched the deep and sincere belief that existed among many archaeologists and historians in the priority of this region in cultural and historical evolution.

According to the *Ex Oriente lux* concept, which was widely popular at that time, most cultural innovations, not to mention civilization itself, came from the East, as the light of the sun does, and historical priority of this region seems to be based on its centuries of experience of unique cultural development.

Theory of Primary Locuses of Nikolay Vavilov and the Problem of Domestic Traits

The next step in the development of migratory views on the spread of agriculture is connected with detection of the earliest manifestations of domestic animals and plants with reconstruction of mechanisms of their introduction into human culture and livelihood practice. During the first half of 20th century, these studies were the cornerstone of the assumption that domesticated forms could appear only in the region where their wild ancestors had been distributed; here the substantial difference between the contemporary and early Holocene environment caused the noncoincidence of modern and prehistoric areals of floral and faunal species, which should be taken into account. Most researchers believe that the most favorable areas for domestication are situated on the margins of different geographic zones, characterized by the junction of landscapes of different sorts. The reason for this belief is that this genre of environment provides broad variability of wild species by way of wide perspectives for hybridization and metisation of related forms of plants and animals as well as for mutation in unstable conditions.

Trying to localize such an environment geographically, Soviet geneticist Nikolay Vavilov referred to the primary loci of plant domestication with arid piedmont areas. In the worldwide context he distinguished seven primary loci of agriculture origin, which he described as rather huge areas where the transition to agricultural mode of life was based on complex cultural plants:

1. East Mediterranean locus, or Fertile Crescent (Iran, Iraq, Israel, Jordan, Syria, Turkey): 9000–7000 BCE, wheat, barley, rye;
2. South Asian locus (Southern China, South-Eastern India, and Southeastern Asia): 7000–5000 BCE, rice, tuberousals;
3. East Asian locus (Mongolia, Amour region, Northern China): 7000–5000 BCE, Chinese millet, beans;
4. Sahara and Sudan: 4000–3000 BCE, pearl millet, sorghum;
5. Guinea and Cameroon: 4000–2000 BCE, yam, beans, oil-bearing palm;
6. Mesoamerican locus (Central and Southern Mexico): 9000–4000 BCE, maize, amaranth, string bean, pumpkin, pepper, garden trees;
7. Andean locus (Colombia, Ecuador, Bolivia): 7000–5000 BCE, potato and sweet potato, manioc, amaranth.

These primary loci of domestic plants introduction are characterized also by introduction of different forms of cattle, most widespread of which were sheep and goat (sometimes hardly differentiated) and bull, which are known in domestic faunal assemblages of 8000–7000 BCE.

Further studies in this field during the second half of 20th century are connected with detailed elaboration of history of certain plants introduction into human culture. It should be stressed that today most of the studies (especially those concerning cereals) disprove Vavilov's basic assumption about the necessity of direct connection in time and space between wild and domestic sorts. A brilliant illustration of this retraction could be found in the history of domestication of wheat, one of the most important cereals in human culture and procurement system. Wheat is traditionally regarded as one of the earliest domestic plants, which was introduced to human culture and productive economy on the territory of Fertile Crescent as early as 9000–7000 BCE. Paleobotanic assemblages of the first settlements of Near East farmers indicate that it was cultivated together with barley even on the same plot of land. From the very beginning early farmers had already explored several types of wheat, most widespread among which were representative of tetraploid (emmer) and diploid (einkorn) wheat.

Identification of the wild ancestors of wheat is complicated by the genetic diversity of its domestic forms, which implies that they were introduced into farming independently. Einkorn (*Triticum monococcum*) is regarded as a relatively more archaic form in comparison with emmer; nevertheless most paleobotanics believe that it was domesticated in the Tavrus region on the basis of its widely distributed wild form a little bit later than emmer. Emmer's wild form, which nowadays is growing freely in southwest Turkey, western Iran, and Armenia, usually is regarded a cytogenetically nonrelated with domestic emmer (*Triticum dicoccum Schrank*). The latter is traditionally connected with the southern coarse-grained race of tetraploid wild emmer, which was primarily domesticated in Palestine and since distributed as a domestic form in the Near East.

Appearance and dissemination of soft wheat in approximately 6000 BCE traditionally is equated with the process of natural hybridization of wild emmer and einkorn and with their wild relative Aegilops that occurred in southern Iran and Transcaucasia. Nanous wheat, which was widely distributed in Caspian region and Caucasus, most probably resulted from natural hybridization of emmer *Triticum dicoccum Schrank* or common (or bread) wheat *Triticum aestivum* and club *Triticum compactum*. It is worthy mentioning that soft wheat would replace emmer only in Roman times. The crucial factor here is the genetic difference of various sorts of wheat—14 chromosomes for einkorn, or small spelt, 28 chromosomes for wild and durum wheat, and 42 chromosomes for soft wheat. The qualitative gap between wild and durum wheat is characterized by an equal number of chromosomes, which could not occur without special genetic changes.

A similar patchy history of domestication and territorial remoteness of wild and domestic sorts is also found for other plants whose productivity was important for prehistoric population all over the ecumene, such as the pea, chick-pea, flax, carrot, and others. These ideas became the starting point for the theory that in the course of extensive migrations of early agriculturalists, early domestic forms of plants were moved away from their wild ancestors within a short period of time.

Spread of Agriculture to Europe: Searches for Sources and Routes

In the light of Nikolay Vavilov's theory of primary locuses of domestication, a transition to productive economy could be viewed as a multicentric process of chronological frames, which are different for different regions and different sorts of plants and animals. At the same time, in global historical context, the earliest traits of domestication of plants were referred to in a strictly defined east Mediterranean zone or the territory of contemporary Iran, Iraq, Israel, Jordan, Syria, and Turkey. In the mid-1960s, Richard Braidwood called this zone the Fertile Crescent and defined more exactly its localization on the border of the Zagros and Tavros mountains with their neighboring steppes.

During the 1960s to 1980s, dissemination of domestic plants and animals over the world was considered a long-lasting process that began in the Fertile Crescent zone since ninth millennium BCE ("effective" village stage, Jarmo culture of the Middle East). Harvest collecting arose from seeds gathering, which preceded it. Earliest evidences of plants domestication are traces at Natufian settlements of Palestine, Shanidar, and Ali Kosh in Iran and Iraq and are dated about 9000–7000 BCE. Paleobotanic assemblages from these settlements indicate that barley *Hordeum distichum L.* and wheat *Triticul dioccum Schrank* were jointly cultivated there with slight domination of barley.

A detailed historical picture of agriculture dissemination from Fertile Crescent to Aegean and Mediterranean region, its exodus to Balkans, and from there to Central and Western Europe in the course of dispersion of Sesklo, Karanovo, Starčevo-Kőrös-Criş, and Linear Pottery (LBK) culture was reconstructed on the basis of archaeological data and updated sequences of radiocarbon dates during the 1970s to 1980s.

At that time it became known that Europe could not be regarded as the mentioned locus, and it became the motivation to connect origin of land cultivation and cattle breeding in Europe with the impact from neighboring territories, first of all, from Asia Minor. In the context of European prehistory, transition to productive economy was developing from its southern regions to the north.

The archaeological context of the starting point of the process of spread of agriculture into Europe traditionally is equated with the exodus of population from southwestern Anatolia to the western and northern Aegean region. Settlements on the Cyprus island (Khirokatia), continental Greece (Sesklo), Macedonia

(Nea Nikomedia), and Crete resulted from this migration are dated by seventh millennium BCE to the beginning of the sixth millennium BCE. From there, early agriculturists moved to Trachea and farther to the north up to the middle Danube region, central Transylvania, and the Balkans, bringing with them the first domestic plants and skills of cattle breeding.

By the first half of sixth millennium BCE, the whole Balkan region was engaged in the process of early agriculturist dispersion displayed in Karanovo in Bulgaria, Starčevo-Kőrös-Criş in former Yugoslavia, Romania, and Hungary. Formation and further durable and gradual expansion of this population on vast territory localized between Phessalia and Tysa basin and between Adriatic and eastern Carpathians and Dnister basin has caused ambiguous progress. On the one hand, it has resulted in neolithization of local Late Mesolithic cultures (such as the case of Lepensky Vir). On the other hand, movements of this population had influenced significantly the process of further agriculture spread into Central and Eastern Europe due to diffusion and relatively rapid dissemination of an agricultural set of know-how (ideas, skills, domestic plants and animals, and techniques of their treatment).

Exodus of Sesklo, Karanovo, and Starcevo-Krish populations from the Balkan region meant overcoming the natural limits of the Mediterranean climatic zone and is traditionally associated with the origin and dissemination of sites attributed to linear pottery (Bandkeramik, LBK) culture. Its core is localized in the Carpathian region, and during the second half of sixth millennium BCE, transmitters of this culture started their dissemination over broad territories of Western and Central Europe. Recent studies based on new series of calibrated radiocarbon data obtained from sites of linear pottery culture show that the duration of spread of the LBK is shorter than the available temporal resolution of the radiocarbon dating. The rate of spread of the initial pottery making is estimated as 1.6 km per year and is comparable to the average rate of spread of the Neolithic in Western and Central Europe.

Since the end of sixth millennium BCE, one can trace coexistence of at least two sorts of secondary locus of productive economy spread over Europe: forest-steppe areal of land cultivation zone and steppe zone of cattle breeding. Population of these zones differed not only by form of food procurement but also by general livelihood systems, social organization, and spiritual sphere. The forest-steppe population have an affinity with the Balkan cultures while the steppe pastoralists have shown their connection with the nomadic populations of the Eurasian steppes.

Early Indo-Europeans as Agriculture Spreaders: Contemporary Theories and Hypothesis

One of the latest developments in the study of migratory or diffusionistic understandings of the spread of agriculture across the world is connected with searches

for the primary homeland of Indo-Europeans peoples. Most hypotheses put forward in this context during the second half of 20th century are based on the same starting point: dissemination of agriculture should be connected with expansion and migration of Middle East inhabitants, which are viewed as absolute pioneers in this field. To the north, European hunter-gatherers adopted agriculture together with appropriate rituals, rites, and spells, which were pronounced using the language of pioneers of land cultivation and ensured through linguistic similarity of Indo-European peoples.

Most advocates of early Indo-Europeans' interpretation of early agriculturists believe that the process of their formation should be viewed in broad chronological frames beginning since the Mesolithic period and transitioning to productive economy. The spread of pre-Indo-European language culture in such a way is usually connected with the dispersion of farming skills, which implies the development of terminology, rites, and customs. It implies sharing of oases or a monocentric theory of transition to land cultivation and cattle breeding, and searches of time and place of Indo-European origin in such a context mean searches of time and place of agriculture origin.

One of the most widespread in contemporary historical and archaeological understandings of pre-Indo-Europeans as early agriculturists was proposed in the late 1980s by Colin Renfrew. Localizing Indo-Europeans in Central and Eastern Anatolia as early as the middle of the eighth millennium BCE, he distinguished 10 diffusions of Indo-Europeans to the adjacent and relatively remote territories (Black Sea steppe region included). Such diffusions caused by necessity ensured facilities for an agricultural mode of life (suitable for farming land) in the situation of considerable growth of population density (up to 50 times in his calculations). He doesn't imply broad human migrations; in Renfrew's understanding it was rather a gradual movement of individuals or their small family groups approximately 1 kilometer per year, which caused a series of local hunter-gatherers population adaptations directed toward acquiring an agricultural mode of life. As a result, it took approximately 1,500 years for agriculture to spread from Anatolia first to Greece, than to the Balkans and southern Italy, Central Europe, and, finally, to Northern Europe as well. No clear archaeological evidence could be found within such a process; it reflects mainly in demographic, economic, and social changes caused by "dominance of agricultural elite."

Similar ideas were expressed almost simultaneously by Soviet researcher Igor Diakonov, who localized Indo-Europeans' homeland in the Balkan and Carpathian regions, indicating that their ancestors could have come from Asia Minor with their domesticated animals and plants. He dated this process to the fifth to fourth millennium BCE.

Russian archaeologist Gerald Matyushin believed that the only common traits for all future Indo-Europeans, which could be traced and proven

archaeologically, are micro-lithic industry and origin of productive economy (land cultivation and cattle breeding). He localized the earliest displays of both these traits in the Zagros Mountains and Southern Caspian region, suggesting that agriculture distribution in Europe should be connected with expansion and migration of Middle East inhabitants to the north (Matyushin, 1986). European hunter-gatherers adopted agriculture together with appropriate rituals, rites, and spells, which were pronounced using the language of pioneers

Traces of two chemicals indicating the earliest evidence of wine were discovered by researchers in this 7,000-year-old pottery jar, from the Zagros Mountains of Iran. (AP Photo/University of Pennsylvania Museum of Archaeology and Anthropology)

of land cultivation, ensuring linguistic similarity of Indo-European peoples. His hypothesis is based on mapping of microlithic technology and productive economy, and temporal and spatial distribution later was proved by linguistic studies of T. Gamkrelidze and Vyacheslav Ivanov. They suggest that the ancestral home of Indo-Europeans was located in the region of Lake Van and Lake Urmia, from where they moved to middle Asia, the northern Caspian region, and southern Ural (Gamkrelidze and Ivanov, 1990).

One more version of so-called agriculturist interpretation of early Indo-Europeans should be mentioned here: the hypothesis about their origin in Central Europe on the territory between Rein, Visla, and the Upper Danube. It was based on the correlation of Indo-European hydronimy to the distribution of population connected with linear pottery culture, funnel beaker culture, globular amphora culture, and corded ware culture. Such understanding of time and place of Indo-European formation shared by G. Kossina, E. Mayer, P. Bosch-Gimpera, G. Devoto, and others was actively discussed during the first half of the 20th century especially by the apologists of Nazi ideas in the human sciences. This discussion had resulted, in particular, by G. Kossina's identification of pre-Germans (or pre-Indo-Germans) with Aryans who were regarded as transmitters of the highest cultural achievements in ancient civilization. This conclusion was broadly used by fascist propaganda as justification of genocide of non-Aryan population practiced in Europe during World War II.

One should confess, nevertheless, that in spite of the existence of fundamental theoretical background and an impressive empirical database that is constantly being updating, the interpretation of early Indo-Europeans as pure agriculturists (as well as their understanding as early nomads) has faced serious objections among archaeologists, cultural anthropologists, and linguists since the very beginning.

Migration: What for? Adoption: Why?

Empirical studies and historical reconstructions of agricultural dissemination across the world traditionally were accompanied by searches for reliable and logical explanations for the causes and backgrounds of this process. In general, it is possible to distinguish two basic directions of such explanations. Proponents of one are concentrating on studies of mechanisms, which forced early agriculturists to move from their homeland, and adherents of the second try to find reasons that make agriculture adoption possible or even indispensable for populations of territories lying outside the primary locus of agriculture origin. An overwhelming majority of the studies in this field fall under the rubric of the first approach, and the basic idea here is that the evictions of groups of population were provoked by disproportion between natural resource availability and procurement requirements. Two basic groups of theories could be distinguished here. On the one hand, most researchers tend to suppose that local population exodus from Asia Minor was caused by demographic factors, first, by rapid growth of early farmer populations and natural restrictions of arable land, which provoked demographic tension and necessity to search for new territories suitable for plant cultivation (Childe 1958; Binford 1968; Renfrew 1989; Braidwood, 1952; and others). Earlier displays of an analogous demographic situation at the Pleistocene-Holocene boundary resulted in the origin of productive forms of economy itself. On the other hand, environmental factors (i.e., changes of natural environment) are regarded as the basic reason for early agriculturists' migration from the Fertile Crescent during the eighth to sixth millennium BCE. In broader context it is viewed as a disruption of equilibrium between nature and human society, which previously also influenced origin land cultivation and cattle breeding (Binford, 1968; Harlan, 1977; Zohary, 1969; and others).

It should be stressed, nevertheless, that in contemporary science the essence of these environmental changes that happened during the 10th to 6th millennium BCE and are understood ambiguously. On the one hand, the population exodus from the Fertile Crescent is explained by general climate aridization in this region, which happened in the course of Holocene and caused a shortage of arable land and crops reduction. This point of view is maintained by numerous palinological and paleobotanic data from archaeological sites of the region under study that highlight paleoenvironmental situations here during Neolithic, Eneolithic, and Early Bronze ages.

On the other hand, recent studies of British and American marine geologists have given new discussion on hypotheses concerning Great Flood in the Black and Mediterranean seas. According to this point of view, spread of agriculture over Europe caused by migrations from the Middle East was provoked by the sea level rise. It was so rapid (according different calculations, 73,000 square km were inundated during 34 years, or over 100 square km during 300 days, or about 60,000 square miles were covered by water, which approached as much as six inches per day) that it was felt by local population as catastrophic (water cover 400 m deep into the coast every day). That's why about 150,000 inhabitants of the Fertile Crescent were forced to displace, and the only suitable place for their procurement system territories were located in northwestern and northeastern directions. In such a way, early agriculturists appeared on the European Mediterranean region and also inhabited the northern Black Sea region (territories of contemporary Bulgaria, Romania, and Ukraine), bringing with them their technology of land cultivation, seeds, and pottery-making skills. In this connection, clarification of reasons that made agriculture adoption possible for populations met with newcomers becomes especially acute. Why didn't they also try to practice new forms of procurement activity instead of fighting with invaders? Why didn't they, in the permanently complicating demographic local population, make a choice to adapt instead of fight for preservation of their cultural and economic identity?

Most contemporary research tends to explain this fact by crisis of traditional hunter-gatherer economy of these territories of and necessity to secure subsistence systems in new ecological situations. This crisis most probably was caused by the new climatic situation of the Middle Holocene as well as by the anthropogenic impact (new demographic situation and critically changed demographic pressure on the foraging territory complicated by nonrational utilization and main hunting species overkills in previous time). In this situation, searches for principally new forms of economic activity were the only possible way to survive, and newcomers proposed to them an already well-elaborated version suitable for introduction.

Searches for Subject of Migration: Human Movements versus Migration of Ideas

It should be stressed that in spite of an obviously vast database and rich arsenal of instrumental techniques, it still remains unclear what the cornerstone of the migrationist version of agriculture spread is—that is, the subject of migration. One can trace at least three approaches to understanding of the essence of such migration. Adherents of the first of them, known now as demic expansion, or wave of advance model suppose direct colonization of Europe by population who moved from the Fertile Crescent (Childe 1958; Ammerman and Cavalli-Sforza 1973; and others). Their reconstructions are based on archaeological data

as well as on scarce but rather illustrative results of paleogenetical studies of the prehistoric population of Europe.

At the same time, the newest results of DNA analysis show no clear signs of large-scale waves of migration to Europe Mediterranean or Middle Eastern groups (Haak et al. 2005). On this basis, a wide variety of "compromise" migrationistic versions were produced. In most cases they imply penetration (infiltration) of small groups of immigrant farmers who came in contact with the aboriginal hunter-gatherers (or horticulturists), creating complicated chains of cultural contacts and conditions for mutual transformations. Highest scientific resonance among such theories of models of elite dominance were proposed by Colin Renfrew (1989), "leapfrogging colonization" by Johao Zilhão, and "individual frontier mobility" by Marec Zvelebil (2000).

Promoters of the third approach to the conceptualization of agriculture spread into Europe as an adoption of "agricultural know-how" (domestic plants and animals, technique and skills of their treatment, rites and customs as well as lingual background of these cultural innovations) by indigenous hunter-gatherers through the diffusion of cultural novelties by means of intermarriages, assimilation, and borrowing (Thomas (2003; Ch. Tilley; Whittle 1996). At the same time, the current database of archaeology, paleoanthropology, paleolinguistic, and natural science (first of all, botany and genetics) provides fundamental background for further development of the relatively balanced point of view expressed in the early 1970s by Ruth Tringham (1971) who considered spread of agriculture into Europe a combination of diffusion and local inventions.

The newest trends in the field of spread of agriculture into Europe are connected with integration of database and research methods of archaeology, radiocarbon dating, genetics, linguistic, and mathematically-based population dynamic modeling. Such studies provided by Pavel Dolukhanov with his colleagues (2005) indicate the necessity to distinguish two historical and cultural processes: one connected with neolitization of Europe and the other displayed in the spread of agriculture into Europe. For the former, they proved a high probability of waves of advance that swept westward through Eastern Europe about 1,500 years earlier than the conventional Near-Eastern one, which resulted in early ceramic sites in western Europe (e.g., La Hoguette-type in northeastern France and western Germany, and Roucadour-type, or epicardial, sites in western Mediterranean and Atlantic France). This earlier wave, dated approximately by 8200 BCE, originated most probably from a vast steppe area stretching between the Lower Volga and the Ural River (e.g., early pottery sites of the Yelshanian culture), spreading from the east via the steppe corridor, resulting in the establishment of the eastern version of the Neolithic in Europe, which implies utilization of ceramics with no apparent traits of agriculture. Agriculture was brought into Europe in the course of the already described later wave originating in the Fertile Crescent of the Near East at about 6700 BCE.

Conclusion

State-of-the art research in the field of diffusionistic or migratory explanation of spread of agriculture over the world indicates that there is no common magistral approach in its understanding in contemporary human, social, and natural sciences. One can trace series of subjects of ambiguous discussions, and most acute and unequivocal among them are definition of migration subject and causes.

Among existing approaches, theories, and hypothesis, one can distinguish two main approaches to conceptualization of background essential for the migratory spread of agriculture. Representatives of the first approach suppose that new forms of human activity, as well as their technological, social, and ritual background alongside with corresponding terminology, have been borrowed by hunter-gatherers from skilled agriculturists in the course of their direct contacts. Causes and mechanisms of such contacts in most cases are conceptualized as migration and as the most probable mechanism of such changes of human lifeways and economy.

Proponents of the second approach believe that the transition to a productive economy could happen only in the case of necessity to change the traditional mode of life, and an insufficient subsistence strategy was comprehended and put into practice by certain groups of local populations. In such a context, even the relocation of bearers of new forms of food procurement into disparate local populations would not inevitably cause the adoption of these forms.

The latest shift of prehistory paradigms toward studies of prehistoric human society in close connection with its environment opens new perspectives for migratory understanding of the process of the spread of agriculture. It implies durable and gradual transition to land cultivation and cattle breeding, which began at certain times and places, as determined by a series of factors and agencies. During the second half of the 20th century this set of circumstances became a subject of discussion in archaeology, cultural anthropology, and prehistory.

The key point of the discussion on diffusive modes of agriculture dispersion is the problem of identification of migration subjects (i.e., establishing who—individuals, local groups, or ethnic groups—or what—ideas, technologies, know-how—was migrating). The only reliable way to verify arguments pro and con every idea and hypothesis is through large-scale paleogenetical studies, organization of which is seriously restricted by scarcity or total absence of paleoanthropological data from certain regions. It should be mentioned that methods of such studies and their reliability are also disputable, so comprehensive observations of possible gene flows remains an acute task for future research projects.

One more crucial problem that is continually discussed in frames of diffusional approach to spread of early agriculture is the definition of the exact time and place where and when wild plants became domesticated. Most probably, this process was durable and was influenced by ancient farmers and cattle

breeders who tried to select the most suitable from among their traits and features. At the same time, some species (mainly floral) have become background for formation of several domesticated plants that showed at the same time no direct genetic connection among them (as was the case of wild and domesticated wheat). It means that mutation mechanisms were equally important in the course of plant and animal domestication as evolutionary ones.

Farming and cattle breeding were regarded only as an additional source of food supply; wild and domesticated species were often represented side-by-side in faunal assemblages and palinological diagrams obtained from cultural layers of Neolithic settlements of Eurasia. This implies that in the course of durable and gradual introduction into productive economy, domesticated species (especially animal) and their wild ancestors could crossbreed many times, and this fact substantially complicates the definition of the domestication process in time and space.

So, at the beginning of the 21st century, the migratory, or diffusive, mode of spread of agriculture from the Fertile Crescent zone remains one of the trustworthy scientific approaches to historical and cultural interpretation of complex process of transition to productive economic forms. Nevertheless, many of its cornerstones are still awaiting comprehensive interdisciplinary studies.

References and Further Reading

Ammerman, A. J., and Cavalli-Sforza, L. L. "A Population Model for the Diffusion of Early Farming in Europe." In C. Renfrew, ed., *The Explanation of Culture Change: Models in Prehistory*. Pittsburgh: University of Pittsburgh Press, 1973.

Andel, T. H., and C. N. Runnels. "The Earliest Farmers in Europe." *Antiquity* 69 (1995): 481–500.

Bennett, J. W. "Ecosystem Effects of Extensive Agriculture." *Annual Review of Anthropology* 2 (1973): 37.

Binford, Lewis R. "Post-Pleistocene Adaptations." In Sally R. Binford and Lewis R. Binford, eds., *New Perspectives in Archaeology*. Chicago: Aldine Publishing, 1968.

Braidwood, Robert J. *The Near East and the Foundations for Civilization: An Essay in Appraisal of the General Evidence*. Eugene: Oregon State System of Higher Education, 1952.

Childe, W. G. *The Prehistory of the European Society*. London: Institute for Archaeology, University of London, 1958.

Clark, J. G. D. "Radiocarbon Dating and the Expansion of Farming Culture from the Near East over Europe." *Proceedings of Prehistoric Society* 31, 4 (1965): 58–73.

Dolukhanov, P., Shukurov, A., Gronenborn, D. Sokoloff, D., Timofeev, V., and Zaitseva, G. "The Chronology of Neolithic dispersal in Central and Eastern Europe." *Journal of Archaeological Science* 32, 10 (2005): 1441–58.

Drew J., Perkins, D., and Daly, P. "Prehistoric Domestication of Animal: Effects on Bone Structure." *Science* 171 (1971): 3968.

Gamkrelidze Thomas V., and V. V. Ivanov. "The Early History of Indo-European Languages." *Scientific American* 262:3 (March 1990), pp. 110–116.

Haak, W., et al. "Ancient DNA from the first European Farmers in 7500-Year-Old Neolithic Sites." *Science* 310 (2005): 1016–18.

Harlan, J. R. "The Origins of Cereal Agriculture in the Old World." In C. A. Reed, ed., *The Origins of Agriculture*. The Hague: Mouton Publishers, 1977.

Harlan, J. R., and J. M. J de Wet. "On the Quality of Evidence for Origin and Dispersal of Cultivated Plants." *Current Anthropology* 14 (1973): 51–64.

Harris, D. R., ed. *The Origins of Spread of Agriculture and Pastoralism in Eurasia*. Washington, DC: Smithsonian Books, 1996.

Higgs, E. S., ed. *Papers in Economic Prehistory: Studies by Members and Associates of the British Academy Major Research Project in the Early History of Agriculture*. Cambridge: Cambridge University Press, 1972.

Mathewson, K. "Cultural Landscape and Ecology III: Foraging/Farming, Food, Festivities." *Progress in Human Geography* 24, 3 (2000).

Matyushin, Gerald. "The Mesolithic and Neolithic in the southern Urals and Central Asia," in Marek Zvelebil (ed.), *Hunters in Transition: Mesolithic Societies of Temperate Eurasia and their Transition to Farming*. New York: Cambridge University Press, 1986.

Perkins, D. J. "The Beginnings of Animal Domestication in the Near East." *American Journal of Archaeology* 77 (1973): 279–82.

Price, T. D. *Europe's First Farmers*. New York: Cambridge University Press, 2000.

Renfrew, Colin. *Archaeology and Language: Puzzle of the Indo-European Origins*. New York: Cambridge University Press, 1987; critically reviewed in *Quarterly Review of Archaeology* 9 (1988) by C. C. Lamberg-Karlovsky and M. Gimbutas.

Refrew, Colin. "They Ride Horses, Don't They?: Mallory on the Indo-Europeans." *Antiquity* 53 (1989): 43–47.

Rindos, D., ed. *Origins of Agriculture: An Evolutionary Perspective*. Orlando, FL.: Academic Press, 1984.

Smith, B. *The Emergence of Agriculture*. New York: Freeman, 1999.

Thomas, J. "Thoughts on the 'Repacked' Neolithic Revolution." *Antiquity* 77 (2003): 75–86.

Tringham, R. *Hunters, Fishers and Farmers of Eastern Europe 6000–3000 BCE.* London: Hutchinson, 1971.

Ucko, P. G., and Dimbleby, G. W., eds. *The Domestication and Exploitation of Plants and Animals.* Chicago: Aldine, 1969.

Vasey, D. E. *An Ecological History of Agriculture, 10,000 B.C.–10,000 A.D.* Ames: Iowa State University Press, 1992.

Whittle, A. *Europe in the Neolithic: The Creation of New Worlds.* New York: Cambridge University Press, 1996.

Yakar, J. "Cultural Interaction between Anatolia and Southeast Europe: An Archaeological Appraisal." In M. Stefanovich, H. Todorova, and H. Hauptmann, eds., *James Harvey Gaul in Memoriam.* Sofia: James Harvey Gaul Foundation, 1998.

Zohary, D. "The Pro-Genitors of Wheat and Barley in Relation to Domestication and Agricultural Dispersal in the Old World." In P. Ucko and G. W. Dimbleby, eds., *The Domestication and Exploitation of Plants and Animals.* Chicago, Aldine, 1969.

Zvelebil, M. "The Social Context of the Agricultural Transition in Europe." In C. Renfrew and K. Boyle, eds., *Archaeogenetics: DNA and the Population Prehistory of Europe.* New York: Cambridge University Press, 2000.

CON

For the longest time in their history, modern humans (representatives of the hominid species *Homo sapiens sapiens*) were foragers who lived by hunting and gathering. According to the reconstructions of human genetics, the gene pool of modern humans originated more than 150,000 years ago. The oldest finds of crania with modern human anatomy were unearthed in central Ethiopia in 2003 and are dated to ca. 160,000 years BP (before present). Until about 11,000 years ago, hunting and gathering comprised the only economic system of subsistence among human beings.

Agriculture, animal husbandry, and nomadism emerged afterward and are, from the standpoint of human history, relatively "recent" forms of subsistence. Although, on a global scale, agriculture is nowadays the most widespread economic system, there are still regions of the world where hunting and gathering have continued as the main sources of food supply. This is true for the Veddah in eastern Sri Lanka and for many of the Brazilian Indians in the rainforests of Amazonia.

Agriculture is one of the great achievements of humankind, and its advent has been described as the "Neolithic revolution." The expression "revolution"

associates certain connotations in our minds, which are, to a great extent, stereotypes of what revolutions are believed to be. A revolution is readily understood as a sudden event that causes a disruption of older traditions and brings about innovations.

In this context, we are confronted with two popular assumptions about agriculture: its allegedly revolutionary breakthrough and its unilateral global spread. Such beliefs are myths of our modern age, and there is good reason to search for the reality behind these myths. We have to ask ourselves:

- Were the beginnings of plant cultivation and the emergence of agrarian lifeways a true revolution?
- Once agriculture had been introduced, did it spread from one region in the world to all the other areas where farming has been practiced?

The reconstruction of reality favors a negative answer to both questions.

Many facts relating to the emergence of agriculture are little known outside the circles of archaeologists and anthropologists, and these need to be discussed in order to draw an overall picture of what happened in prehistory. The factors involved in the development and spread of agriculture must be mapped out to facilitate a substantive answer to the crucial issue of the widespread adoption of a complex set of new technologies.

Agriculture is a blanket term for several forms of economic subsistence that are often interconnected. Those who practice agriculture also keep domesticated animals like pigs, sheep, goats, and cattle. The production of grain as well as meat requires storage capacities (i.e., vessels to store food for later consumption). Therefore, the making of pottery developed as a special technology parallel to the other technologies. In view of the network of technological skills that are involved, it becomes evident that agriculture includes an entire package of diverse activities and technologies.

Experts speak about the "agrarian (or agricultural) package." In such contexts where the emergence of agriculture is referred to—that is, to the domains of plant cultivation, animal husbandry, pottery, and related items—the agrarian package is defined as "the sum of traits that appear repeatedly in the Neolithic assemblages of SW Asia, Anatolia and SE Europe" (Çilingiroglu 2005: 3).

For many years, it was believed that the combination of technologies in the agrarian package is a prototype for all agrarian lifeways in the world. This belief made it easy to think of a spread of this prototype from one region to others. However, the reality of what happened in the Neolithic age is more complex and shows regional variants of a transition from foraging by prehistoric hunter-gatherers to various forms of sustained subsistence.

Recent findings in the archaeology of Eastern Europe call for a revision of the idea of one single package of new technologies. The Neolithic revolution

Package	Technologies	Lifeways	First Appearance
Agrarian	Plant cultivation + animal husbandry + pottery + weaving	Sedentary	Near East
Pastoral	Animal husbandry/ herding + pottery + weaving; no plant cultivation	Semisedentary or nomadic	Eastern Europe (steppe zone of the Pontic-Caspian region, extending between the Caspian Sea and the Black Sea; southern Russia)

produced at least two foundational packages of technological innovations, as shown in the table above.

What makes these new insights about a second package sensational is the evidence that this package did not emerge as an offshoot of the agrarian package but as an independent form of subsistence. It was believed earlier that pastoralism and nomadism originated in regions where people had established agricultural communities, but the cultivation of plants was so scanty that they gave up their sedentary lifestyle to focus on herding as the main form of subsistence.

The archaeological record of pastoralist cultures in the steppe zone of southern Russia shows that herding developed in the region long before small-scale agriculture was introduced. The early variant of pastoralism there did not originate as an offshoot of agriculture as did nomadism in many parts of the world in later periods. Nevertheless, the pastoralists of the Russian steppe produced pottery as early as their contemporaries in southeastern Europe. Their oldest pots date to ca. 7000 BCE.

These insights about the diversity of socioeconomic patterns that resulted from the transition to sustained subsistence are instructive because they illustrate that the human mind is very versatile when it comes to adjusting to varying ecological conditions and to exploiting available resources in an optimal way. In the Near East, local resources allowed the domestication of both plants and animals and the fabrication of pottery, while, in the steppe zone of Eastern Europe, plant cultivation remained unsuccessful, and herding was rightly promoted as the most suitable form of subsistence for that ecological environment.

The awareness of this flexibility of the human adaptive skills to make alternative choices for optimal solutions has a bearing on the way to approach the issue of the spread of Neolithic technologies. Whatever answer we seek to the question of spread we have to specify which package we refer to.

The Long Trail to Agriculture as a System
of Economic Subsistence

The diet of the people who lived during the maximum glaciation of the last Ice Age, some 20,000 years ago, already included protein-rich seeds of several wild plants from the Near East, among them species of wild grass and emmer. Due to their taste and availability, hunter-gatherers began to experiment with them at a later period. Although wild seeds played an ever increasing role in nutrition, the annual period of their availability was always limited.

Human inventiveness was certainly spurred and the human mind was challenged to collect seeds in greater quantity once they had ripened and to devise new means to prolong the span of time for consumption. It can be conjectured that the early experiments with wild seeds were accompanied by simultaneous efforts to improve the capacities of their storage.

The emergence of agriculture as sustained subsistence was neither accidental nor was it purely intentional. Instead, it resulted from the dynamic interplay of various factor, both external (relating to the conditions of environmental ecology) as well as internal (posing a challenge to human adaptive strategies).

When we speak of agriculture we mean full-scale plant cultivation. The beginnings of this phase of food production date to ca. 8500 BCE, and the earliest evidence for it comes from the Near East, from the settlement of Jericho on the west bank in the valley of the river Jordan. Before that time, human beings had experimented with the seeds of wild plants. The path to agriculture evolved step by step over thousands of years. Experimenting with the seeds of wild plants is a necessary precondition for intentional plant cultivation. It is therefore only natural that, in the oldest settlements of local agrarian populations that were established in different parts of the world, one finds a similar trajectory, leading from the exploitation of wild seeds to their intentional cultivation.

The span of time from the first experiments with wild seeds to full-scale agriculture extends over 2.5 thousand years, from ca. 11,000 to ca. 8500 BCE. Is it reasonable to describe the step-by-step evolution that is involved in terms of a revolution? No, it is not. The view of the modern observer who looks back into history tends to condense the proportions of absolute time, and longer periods may look, from the distance, more compacted than they really were.

The experiences of the hunter-gatherers with wild seed plants posed the challenge to remain in the area of their natural spread to follow the ripening and to engage in harvesting at the right time. This process of a seasonal presence of bands of hunter-gatherers in areas of optimal food supply furthered a development that would become a major thrust in human socioeconomic evolution, the trend to become sedentary.

Sedentism is a necessary precondition for the observation of the life cycle of wild plants and their seed production over a long duration. Therefore, sedentism
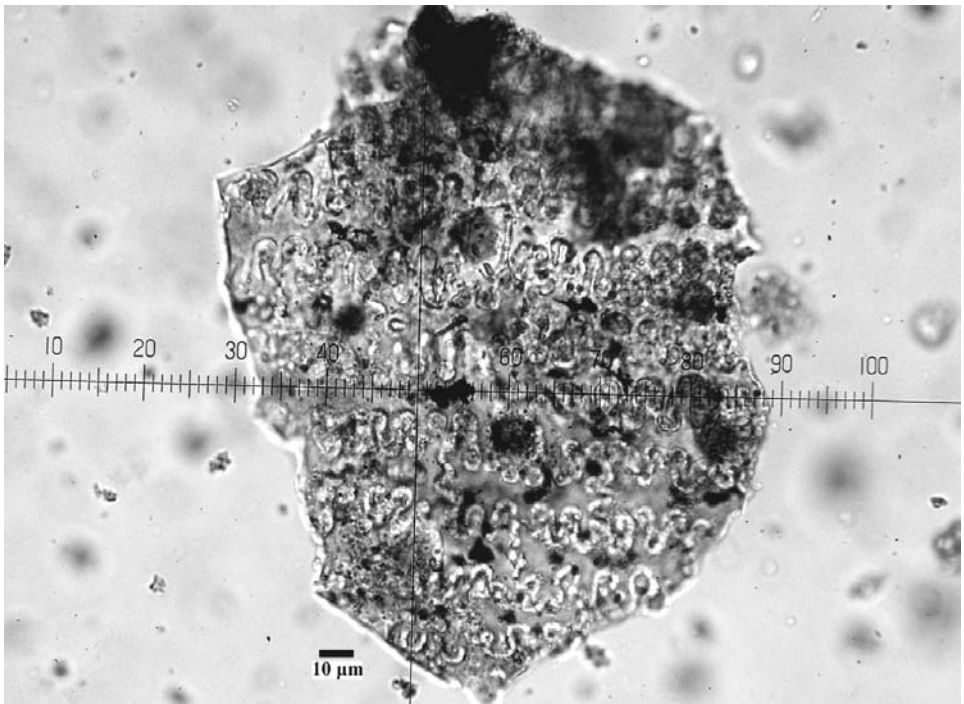
Multi-cell phytolith from the husk of wild emmer wheat (Triticum dicoccoides). Excavated from the Late Natufian site of Raqefet Cave in Israel (12000–9700 BCE). (Arlene M. Rosen/site excavation by Dani Nadel)

always precedes and predates the socioeconomy of plant cultivation. Sedentary lifeways enhance the transition to food production, in the form of horticulture, on a small scale and, as agriculture, on a larger scale.

Plant cultivation was the final stage in a series of developments that evolved in the following order:

- Introducing the seeds of wild plants into the diet of hunter-gatherers;
- Exploiting wild seeds—where available—as food stuff by collecting them in larger quantities;
- Relying to a greater extent on wild seeds in refugia where hunter-gatherers become sedentary;
- Intentional experimenting with wild seeds;
- Domestication of wild plants (i.e., selecting the seeds of particular species for planting);
- Small-scale plant cultivation (horticulture);
- Full-fledged plant cultivation (agriculture).

On inspection of the activities involved in the evolutionary chain of human ecology as described above, one recognizes two basic capacities of human beings

that are operating. One is the network of adaptive skills that enables human beings to adjust to changing ecological conditions in their natural habitat. The other is human intentionality, the mental driving force that is a manifestation of human higher-order consciousness.

Human intentionality furthers the sense of curiosity and spurs the drive to explore and experiment. The evolutionary chain has been reconstructed from the archaeological record of early transitions to sedentism and plant cultivation. Where and when can we find the earliest evidence for ecological conditions furthering such transitions?

Neolithic Transitions

The Emergence of the Agrarian Package in the Near East

The earliest signs of a trend toward sedentism are found in the ancient Near East, in the context of ecological changes that affected human living conditions. When the climate became colder during the Younger Dryas phase (in the course of the 12th millennium BCE), this affected the availability of seeds of wild grass and other food crops. These food resources had been an important asset of nutrition for the hunter-gatherers of the region. In response to the ecological change, humans retreated to areas where wild grass was still available. In the vegetational refugia, humans adapted to local ecology and exploited the food resources intensively. The chain reaction of events during that phase of a shift from mobility to sedentism illustrates that the impulse for that development was sparked by ecological changes, which, in turn, induced responses of human intentionality to cope with those changes.

Sedentism is a stable feature in any agrarian society. In order to reach the stage of full-fledged agriculture, sedentary populations in the ancient Near East experienced a step-by-step transition:

- Sedentism of the Natufian people in the 12th and 11th millennia BCE (intensive exploitation of seeds of wild plants; beginnings of an intentional planting of wild seeds; no domesticated animals); Natufian settlements have been found at sites in Israel, Jordan, and Lebanon;
- Sedentism of the early monument builders (at Göbekli Tepe, Çayönü, Nevali Çori, and other sites in eastern Turkey) in the 10th and 9th millennia BCE (intensive exploitation of wild plants; intentional planting of wild seeds; no domesticated animals);
- Sequence of sedentism of the early inhabitants of Jericho (ca. 10,000–8500 BCE: sedentary hunter-gatherers; ca. 8500–7300 BCE: domesticated plants; ca. 7300–6000 BCE: domesticated animals—sheep and goats).

The history of sedentism and plant cultivation during the following millennia is the history of the spread of full-fledged agriculture and animal husbandry

in the area that has become known as the Fertile Crescent, extending from Egypt in the south over the Near East to Anatolia in the north and Mesopotamia in the east. Those populations that founded the early civilizations of the Ancient Orient in the fourth millennium BCE (i.e., Egypt, the city–states of ancient Sumer and the kingdom of Elam in the Zagros Mountains) were agrarian communities.

The Emergence of the Pastoralist Package in the Steppe Zone of Eastern Europe

Toward the end of the Ice Age (about the mid-10th millennium BCE), the zone north of the Black Sea experienced a stage of progressive desiccation. The second spike of the water from the melting ice (beginning about 9400 BCE) never reached the ancient Euxine Lake, the predecessor of the Black Sea, and aridification of the area north of the freshwater lake proceeded rapidly. The ecological preconditions for the beginnings of pastoralism among the ancient populations on the eastern margins of Europe are found in the forest-steppe zone during that period. The ongoing process of desiccation in the northern Pontic zone caused an extension of the steppe zone in the south and a receding of the forest-steppe belt to the north.

These environmental changes had long-term repercussions on human ecology. Gradually, the development in the south shifted economically, culturally, and linguistically. This was a prolonged process that might have taken more than a millennium to unfold. The people in the southern steppe zone that have been identified as Indo-Europeans experienced a socioeconomic transition from foraging to herding. Since the climatic effects of desiccation enhanced the transition to pastoralism in the Pontic steppe zone, the development there in the communities of Indo-European stock detached itself from the former socioeconomic, cultural, and linguistic basis of convergence with the people farther north, the Uralians, resulting in the formation of an epicenter of old Indo-European culture.

The linguistic and archaeological evidence speaks in favor of pastoralism as having developed independently from farming in the steppe zone of southern Russia. This means that the agrarian package that reached Europe from Asia (see the next section), spread from southern to central Europe and, later, farther east, had no part in processes of transition that took place in the steppe zone. The transition from a foraging to a pastoralist economy was accompanied by changes in lifestyles. Historical linguistics have reconstructed an old layer of common lexical roots for the domain of pastoralism in Indo-European languages.

This terminology forms part of the core vocabulary of the Indo-European protolanguage, the reconstructed common basis from which all Indo-European languages derive. The old layer of terms for herding is widespread across the entire range of Indo-European languages and has been preserved in most of them up to the present. The wide spread of such terminology is an indication of the general importance of this vocabulary for the early Indo-Europeans. The persistence of

the old diagnostic terms of herding, especially in the eastern Indo-European languages, points toward the steppe zone as the area of origin for pastoralism.

The Collision of Worlds—Agriculturalists and Pastoralists in Conflict in Prehistoric Europe

It is obvious that agriculture spread within the wider region of the Fertile Crescent because the people who populated the areas lived in contact with neighboring communities and shared common socioeconomic knowledge with their neighbors. It is also obvious that agriculture spread from Anatolia to Europe. The earliest evidence for plant cultivation is found in southeastern Europe, dating to around 7000 BCE. It is still a matter of dispute whether the agrarian package was brought to Europe by migrants from western Turkey or reached the indigenous Europeans via idea diffusion. The facts from modern archaeology point in the direction of idea diffusion and acculturation because there is no evidence for a major migratory movement from Anatolia to the west at the time of an early appearance of plant cultivation in Europe.

During the seventh and sixth millennia BCE, the indigenous population in southeastern Europe adopted the agrarian package, and agriculture spread as far as Ukraine. On the western periphery of the steppe, where the terrain that was frequented by the pastoralists and their herds bordered the area of the arable land (i.e., in southern Ukraine), the boundaries between the two economic systems of pastoralism (the eastern tradition) and of agriculture (the western tradition) began to float soon after ca. 5000 BCE. The initial contacts between pastoralists and agriculturalists may have been peaceful, but things changed when the socioeconomic sphere of the agriculturalists experienced its expansion to the east and new agrarian settlements were established in areas formerly frequented by pastoralists.

Direct consequences of this expansion were an infringement of the movements of the pastoralists and a reduction of their resources, the pastures that turned into fields. And yet there were other consequences that had an even stronger impact on the sustainability of pastoralism in the contact region. The sharing of pastures among pastoralists was organized by rigid regulations, and the prehistoric expansion of agrarian settlements into the steppe zone caused more disturbance than an areal occupation of former pastures.

The consequences of this expansion culminated in a shakeup of the nomadic socioeconomic system and in a threat to the accessibility of resources. The magnitude of this threat might have been felt differently in the regional groups of pastoralists, but, in principle, the world of the agriculturalists exposed itself to the herders as harmful to their socioeconomic sustainability.

Given these unfavorable conditions, the expansion of the agrarian system of subsistence to the east caused increased friction between agriculturalists and

pastoralists, stirring up ever more competition over the exploitation of the terrain. During the first half of the fifth millennium BCE, under the impression of a growing socioeconomic stress in the local communities, there are signs of clashes and even warfare between the western agriculturalists and the steppe people, as evidenced by layers of ashes and an increased number of arrowheads in the archaeological record of the easternmost settlements.

By the middle of the fifth millennium BCE, ever more settlements on the eastern periphery were fortified, and the frequency of arrowheads in the archaeological record increases. Arguably, the migrations of the steppe people find their ultimate motivation in elementary counterreactions to those scenarios of unrest. It is reasonable to assert that the early pastoralists in the Pontic-Caspian zone were challenged by the advance of agricultural practices into the region and their reactions to this "intruding" factor triggered a chain reaction: the Kurgan migrations of the fifth and fourth millennia BCE that brought Indo-European pastoralists from the steppe zone to the west. Archaeologist Marija Gimbutas (1974, 1991) coined the overarching term "Kurgan migrations" to define the movement of the steppe people and to identify the bearers of the earliest recognizable Indo-European culture, that of the people who built huge burial mounds, called *kurgan* (a word of Turkic origin).

The Kurgan migrations unfolded in three consecutive waves, between ca. 4400 BCE (beginning of the first wave) and ca. 3000 BCE (end of the third wave). The process of Indo-Europeanization of the northwestern and western Pontic region was not necessarily the result of massive population movements. The cultural and linguistic changes most probably resulted from the exertion of control of a ruling elite over people and territory either by intermarriage into families of local dignitaries or by assuming power through conquest.

As a rule, the culture of the elite dominates and its language is more prestigious than that of the local population, eventually resulting in assimilation and language shift of the latter. In this process, elements of the local language are absorbed as a substratum by the dominating language. The same holds true for cultural patterns, like the survival of cults of female divinities among the ancient Indo-European peoples in southeastern Europe (i.e., Thracians and Illyrians).

In the northwestern Pontic region, the incursions of the steppe people produced permanent patterns of change. Judging from richly equipped graves, a new social elite makes its appearance at Durankulak (northeastern Bulgaria) around 4600 BCE and, a hundred years later, the tradition of burials also changes at Varna. There, insignia such as a horse-headed scepters and other ceremonial items of political power provide evidence "of the spread of steppe tribes from the east to the west and in the 'Kurgan' model of Indo-European origins is seen to reflect the first wave of Indo-Europeans from their homeland in the steppe-lands of the Ukraine and south Russia" (Mallory and Adams 1997: 557).

The Indo-European pastoralists might have imposed their language and culture on the local population that they ruled, but they changed their own life-ways, becoming agriculturalists themselves within a few generations.

Varieties of the Agrarian Package in Other Parts the World

The identification of the direction of the spread of agrarian technology from the Fertile Crescent to Europe to northern Africa to the Iranian Plateau and the Indian subcontinent is conclusive with the sociocultural interconnections between those regions, and these interconnections show in the archaeological record. However, assuming spread from the Fertile Crescent beyond those areas mentioned here would be based on pure speculation.

In fact, imagining rice cultivation in eastern and southeastern Asia to have originated as a result of influence from the west is more than improbable, it must seem odd. The technology of planting rice in wetlands is very different from cultivating cereals on dry land because the technologies differ greatly. Rice cultivation in eastern Asia (i.e., central China) originated around 7000 BCE upon foundations of locally specific experiences and experiments with wild rice.

It is more than obvious that plant cultivation in the Americas could not have been inspired by idea diffusion from western Asia. The domestication of bottle gourd emerged in central Mexico already around 9000 BCE. This is a local development as much as the cultivation of beans since ca. 8500 BCE and corn (maize) since ca. 7500 BCE. Since, in pre-Columbian America, the beginnings of horticulture are more or less contemporaneous with the transition in the Fertile Crescent, the possibility of a transfer of technological know-how from western Asia to the Americas can be reasonably excluded.

There are other criteria that would make Asian-American interactions in matters of plant cultivation unlikely even if a time lapse were involved. There is the factor of seclusion of the early epicenters of sedentism and domestication of plants because "agriculture in America began in areas like central Mexico and the western part of South America (mostly Ecuador and Peru), which were to some extent unique or isolated" (Cavalli-Sforza et al. 1994: 308). Thus, plant cultivation on the Mexican plateau and in the Andean region had to rely on local experiences with climatic conditions and with specific qualities of the soil suitable for certain species.

The Facts Speak in Favor of Local Trajectories of a Transition to Agriculture

In eastern Asia, agriculture emerges with some time lapse, that is some time later than its early beginnings in the Fertile Crescent. This has prompted the popular notion that the agrarian package, with its related technology, knowledge,

---

Maize Domestication in Ancient Mexico

One of the most convincing pieces of evidence for the argument that agriculture, the domestication and production of crops, rose at different places around the world independently is the archaeological discovery of pollens and patterns of agricultural life dating back millennia. A 2001 article in *Science* magazine presented evidence that agricultural behavior patterns can be detected as far back as 5100 BCE in the Grijalva River delta of Mexico. Along the Gulf of Mexico in the region of Tabasco, archaeologists have found evidence of forest clearing as well as pollen from cultivated Zea, which was an early form of domesticated maize. In addition, pollen from domesticated Manihot has been found dating to 4600 BCE, and evidence that other crops, such as sunflowers and cotton, were being produced dates back to 2500 BCE.

Source: Pope, Kevin O., et al. "Origin and Environmental Setting of Ancient Agriculture in the Lowlands of Mesoamerica." *Science* 292, 5520 (May 18, 2001): 1370–73.

---

and lifeways, would have been allegedly transferred—either via direct contacts (e.g., via trade) or via idea diffusion—to all other regions where agriculture began to be practiced. This notion of a monogenesis, that is, the idea of plant cultivation as an invention that was made only once in human history, was coined more than a hundred years ago, at a time when the information available about the early phases in the development of agriculture was fairly limited and the pool of data for each region had few specifics. Furthermore, that was a time when the biology of human social and cultural evolution was still in its infancy and the dynamics of human adaptive skills had not been studied in more detail.

The notion of the monogenesis (from Greek *mono-* for only, one only, alone plus *genesis* for origin, source, creation) of agriculture as a human invention is a typical product of the Euro-American way of thinking. In our modern era, the mind is set to competitive patterns of human inventiveness. Widespread are beliefs that, once an invention is made and its usefulness is demonstrated, then it is exported and prone to spread throughout the world. Similar statements regarding the unilateral expansion of plant cultivation have been made about various economic or cultural innovations in human history, for example, for the tradition of rock art, the making of pottery, weaving, metal-working, writing, and other skills.

It is well known that many inventions were made several times at different places and at different times. The skills of how to light a fire unfolded in human communities around the world without people migrating and taking the knowledge from one region to another. Modern science, archaeology, anthropology, and culture studies in particular have produced evidence that all these technologies originated spontaneously in various regions of the world independently. The Native Americans developed their skills of pottery making on their own,

and they did not need any inspiration from Asia to activate their artistic abilities. The ancient Chinese developed their writing technology without any outside inspiration. Among the criteria that speak for an independent origin of Chinese characters is the fact that the script was used for purposes (i.e., divination) that were unknown in Mesopotamia with its seminal writing technology the cuneiform script.

In the past two decades or so, many findings and insights have been produced by various disciplines of the natural sciences and the humanities that speak in favor of an independent development of agrarian technology in those parts of the world that were not interconnected by trade and where there is no indication of migrations of people or cultural exchange.

The historical facts about sedentism and full-fledged agrarian lifeways, which is the reality of human socioeconomic history of the transition from the Paleolithic to the Neolithic age, do not favor the idea of a revolutionary advent of agriculture. Moreover, reality speaks for a spontaneous and independent drive toward sedentism and plant cultivation in different parts of the world and against the assumption that the advent of agriculture was a unique innovation that happened only once in human history. The amount of technological know-how and the variety of technologies by which local agrarian packages distinguish themselves are so diverse that even the theoretical idea of a global diffusion of basic skills spreading from one single region seems unrealistic.

Besides, human adaptive skills and intentionality have been operating in all parts of the world. Human beings in western Asia, in eastern Asia, and in the Americas made their own experiences with ecological conditions, which, independent from one another, culminated in similar drives toward sedentism, the cultivation of plants, and the domestication of animals. From the early epicenters, the technologies and lifeways of the "agrarian package" spread, in secondary movements, to neighboring regions. Secondary diffusions are, for example, rice cultivation in Southeast Asia, which was introduced from China, the early appearance of agriculture in the southwestern United States, which resulted from the spread of the agrarian package from central Mexico, or the spread of plant cultivation in western Africa, which was inspired from agriculture practiced in the Nile Valley.

Thus, the reality of secondary spread and diffusion does not stand in contradiction to the concept of multiple origins of agriculture, provided one keeps in mind how their dynamic phases are related to the frame of absolute chronology.

References and Further Reading

Bellwood, Peter, and Colin Renfrew, eds. *Examining the Farming/Language Dispersal Hypothesis*. Cambridge: McDonald Institute for Archaeological Research, 2002.

Boyle, Katie, Colin Renfrew, and Marsha Levine, eds. *Ancient Interactions: East and West in Eurasia*. Cambridge: McDonald Institute for Archaeological Research, 2002.

Cavalli-Sforza, L. Luca, Paolo Menozzi, and Alberto Piazza. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press, 1994.

Çilingiroglu, Çiler. "The Concept of 'Neolithic Package': Considering Its Meaning and Applicability." *Documenta Praehistorica* 32 (2005): 1–13.

Gimbutas, Marija. *The Gods and Goddesses of Old Europe, 7000–3500 B.C. Myths, Legends, Cult Images*. London: Thames & Hudson, 1974 [2nd ed. 1982 as *The Goddesses and Gods of Old Europe, 6500–3500 B.C.*].

Gimbutas, Marija. *The Civilization of the Goddess: The World of Old Europe*. San Francisco: Harper, 1991.

Haarmann, Harald. *Foundations of Culture—Knowledge-Construction, Belief Systems and Worldview in Their Dynamic Interplay*. New York: Peter Lang, 2007.

Haarmann, Harald. "Indo-Europeanization—The Seven Dimensions in the Study of a Never-Ending Process." *Documenta Praehistorica* 34 (2007): 155–175.

Harris, David R., ed. *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. London: UCL Press, 1996.

Mallory, J. P., and D. Q. Adams, eds. *Encyclopedia of Indo-European Culture*. Chicago: Fitzroy Dearborn, 1997.

3

The Great Flood referred to in the Book of Noah and in Gilgamesh resulted from the flooding of the Black Sea by an influx of higher-level water from the Mediterranean via the Dardenelles and Bosporus.

PRO   Harald Haarmann
CON   John Lee

PRO

Stories about great floods are found in many cultures of the world, and there is growing scientific evidence that such natural catastrophes have happened in various parts of the world at different times. Natural scientists have found proof that some prehistoric floods may have been caused by volcanic eruptions or earthquakes that triggered massive landslides in coastal areas, producing tsunamis (i.e., tidal waves) of great magnitude and high destructive potential. For several regions, the connection between local flood myths and geologic realities of flood events is well documented.

In the Indian flood myth, written in the Sanskrit epic "Mahabharata," Manu—the equivalent of Noah—survives the catastrophe. When the flood account is related to the archaeological record of northwestern India, one finds a match in the river valleys of the Indus and its tributaries. There, in the urban environment of the pre-Aryan cities, the ancient Indus civilization (Harappan civilization, respectively) flourished during the third and early second millennia BCE. The decline of this civilization is ultimately linked to catastrophic flooding of the area. "In time, layer after layer of Indus mud, possibly wind-blown as well as water-borne, choked the streets, rotted the timbers, and piled high above the rooftops. . . . On top of the cities, now consigned to oblivion beneath tons of alluvium, other people grazed their goats, sowed their seeds and spun their myths. A great civilisation was lost to memory" (Keay 2000: 5).

In another region, flood events can be regularly observed and there can be no wondering why the local people have entertained flood myths for many generations. The landscape of the Amazon region in northern Brazil, with its vast extension, changes according to the seasons of the year. In spring, many areas are flooded by the waters of the inundating Amazon and its tributaries. This river system receives water supply from the Andes, and the flow of water varies

51

according to the amount of snow that melts toward the end of winter. The water level of the Amazon River may rise between 3 to 10 meters in spring, resulting in the massive flooding of areas that lie dry in summer. In the animistic world of the Amazon Indians, flood events are mystified and associated with the doings of powerful spirits.

The specific flood of interest here (i.e., the Black Sea flood) has been associated with Noah's flood in the biblical account and is assumed to have occurred in the Black Sea Pontic region in prehistoric times. The Black Sea was not always connected to the Mediterranean. During the Ice Age, Europe and Asia were connected by a land bridge, a hilly terrain that allowed easy passage of humans and animals in either direction. On the northern side of this land bridge extended a freshwater lake (i.e., the former Euxine Lake). The land bridge was broken, the lake and the surrounding areas were flooded by marine water from the south, and the Black Sea, as we know it today, originated.

How this exactly happened is still a matter of much debate. Was the flood event catastrophic or did the marine water spill over the once existing land bridge gradually? Are the shorelines of the Black Sea the result of one major flood event, or are subsequent sea-level changes also responsible for their profile? How often in the history of the Black Sea did sea-level changes occur?

According to the geologic findings that William Ryan and Walter Pitman presented in their 1998 study *Noah's Flood*, the flood was one catastrophic event that changed the ecology of the region, once and for all. The information about the Great Flood that Ryan and Pitman presented has stirred up much criticism, at the same time it stimulated intensified research on phenomena relating to catastrophic floods in world history.

What's in a Flood Myth?

> *The Flood roared like a bull,*
> *Like a wild ass screaming the winds [howled]*
> Atrahasis III, OBV iii; quoted after Dalley 1998: 31

Of all the stories about the Great Flood, Noah's flood, as reported in the biblical account (Genesis 6:9–9:17), is the most widely known of all those tales. When Charles Leonard Woolley excavated the ancient city of Ur in Mesopotamia in the 1920s and dug into 10 continuous feet of waterborne silt with no trace of human artifacts, he thought that he had found evidence testifying to Noah's flood. For decades, archaeologists and the wider public alike were convinced of Woolley's claim that the traces of a flood he had found in southern Mesopotamia corroborated the biblical report.

Even before the discovery of the Black Sea disaster, experts had become aware of numerous flaws in Woolley's hypothesis. The flood at Ur corresponded

---

A Businessman Finds Noah's Ark, or At Least Some Publicity

In April 2004, Daniel McGivern a professed Christian businessman from Honolulu, made a new claim on an old theme: he said he had located Noah's Ark on the slopes of Turkey's Mount Ararat in satellite photos and was planning an expedition that summer to the site to take photographs and establish proof of the veracity of the biblical flood myth found in the book of Genesis. Calling it "the greatest event since the resurrection of Christ," McGivern said that he would fund the expedition to the tune of $900,000, and needed only the Turkish government's approval to make the trip. The announcement made such a splash that *National Geographic* even covered it.

Unfortunately for McGivern, the Turkish approval never came. The mountain is located in a military zone, and permission to climb is rarely given. Of course, the involvement of Ahmet Ali Arslan, who was instrumental in the making of CBS's sensational 1993 *The Incredible Discovery of Noah's Ark*, which turned out to be a deliberate hoax, did not help matters.

Ark sightings have become commonplace, especially since the advent of satellite technology, but most serious researchers discount any possibility that wood from a craft built thousands of years ago would survive to this day. The Turkish government has even claimed that they have evidence of groups scaling the mountain to place wood there in order to "discover" it later. After centuries of people looking for the ark, claiming to find it, and being disproven, the only answer that is conclusive is that the story of the ark is what it has always been—a matter of faith.

Sources: Lovgren, Stefan. "Noah's Ark Quest Dead in Water—Was It a Stunt?" *National Geographic News*, September 20, 2004. http://news.nationalgeographic.com/news/2004/09/0920_040920_noahs_ark.html; Mayell, Hillary. "Noah's Ark Found? Turkey Expedition Planned for Summer." *National Geographic News*, April 27, 2004. http://news.nationalgeographic.com/news/2004/04/0427_040427_noahsark.html

---

to a sequence of inundations rather than to one great flood in southern Mesopotamia. Moreover, the archaeological evidence for Noah's flood was too scanty to have impressed people's cultural memory so deeply in such a wide area. In the absence of more substantial clues, the existence of flood tales in Syria and Palestine was explained as reflecting the import of narratives from a common southern source.

There are also inconsistencies in the mythic tradition relating to the flood. In the oldest Sumerian myth of Atrahasis (the Sumerian equivalent of the biblical Noah), the hero is introduced as a citizen of Shuruppak. Evidence for the flood there is younger and weaker than the strata of silt deposits at Ur. Therefore, a flood that occurred in the Shuruppak area was minor when compared with the flood in the region of Ur. If Atrahasis is the hero with a the great flood, why is his story not related with the old city of the kings at Ur? Why would the flood in the Sumerian myth be described as "roaring"? The seasonal inundations of
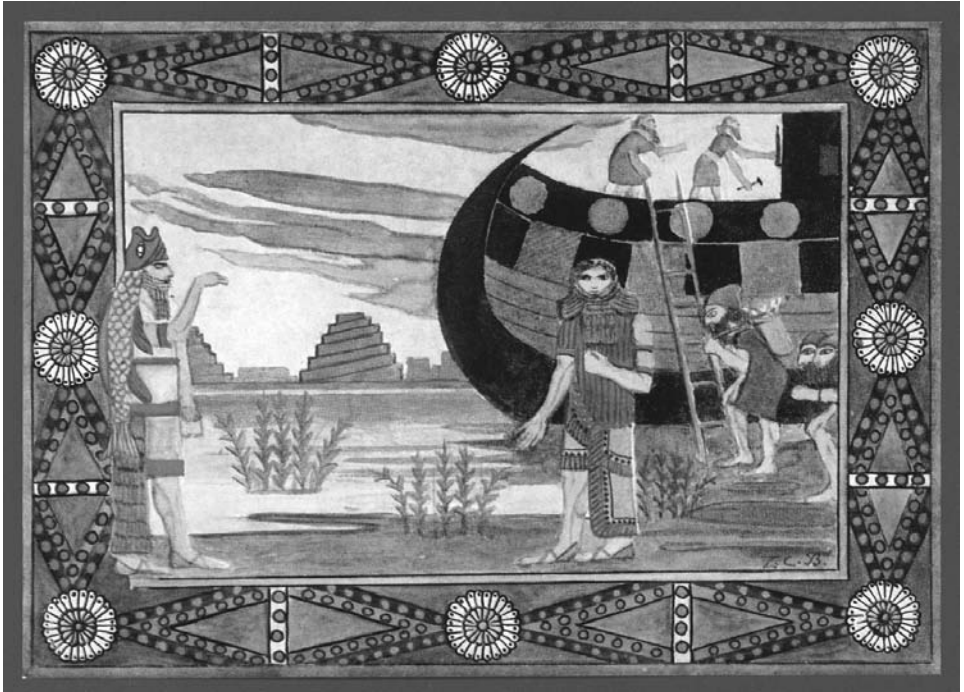
In this illustration of the Mesopotamian flood myth, when the gods punish the wicked city of Shurripak with a Deluge, they tell Utanapishtim to build a boat and take every kind of living thing aboard, about 2000 BCE. (Mary Evans Picture Library/The Image Works)

the rivers in southern Mesopotamia were a rather silent event, with the waters rising slowly.

It is hard to understand why the successive floods in southern Mesopotamia would have inspired local people to create a myth of one great flood that destroyed an older world and gave rise to a new world order. In the predeluge world of the Sumerian flood myths, the gods were near and the effects of their doing were felt closely, sages acted as culture heroes, and brave men could still obtain immortality. In the postdeluge world, even the sages became mortal and there was no longer immortality for any of the living.

In fact, the narrative impact of the ancient flood myths can readily be associated with the long-lasting after-effects of a traumatic experience that shaped the cultural memory of people. When the land bridge between Europe and Asia broke, the echoes of the dramatic opening of the Bosporus Strait were perpetuated through many generations. In view of the migratory thrusts into different directions away from the inundated shores of the former freshwater lake, myths relating to the Great Flood emerged spontaneously among the local migrants, and the oral narratives were conveyed to their offspring. There can be no wonder that flood myths also persisted among those who later lived along river basins such as the Sumerians.

Flood myths are found in the narrative traditions of Mesopotamia, the Near East, and southeastern Europe. The best known Greek flood myth is perhaps the story of Deucalion, son of Prometheus, and Pyrrha, daughter of Epimetheus, who survive a flood that devastates Thessaly. The parallelisms found in the stories are striking. Of particular interest is the lake-burst motif that "could fit several locations both in Greece and Turkey, but the mother of the legend could have been super floods surging either way between the Black Sea, the Sea of Marmara and the Aegean" (Oppenheimer 1998: 261).

If the European and Mesopotamian traditions of flood myths have common roots, there is a high probability that those roots are linked to the natural disaster that happened ca. 6700 BCE. This would represent an incredibly long trajectory of cultural history comprising many millennia in which the story of the flood was imprinted in people's memories from one generation to the next.

The narrative potential of that event has assumed an intercultural magnitude with repercussions even in regions where there have been no actual floods. An example of this is the pre-Christian sea-flood motif found in certain carols and dirges of Romanian folklore. It is noteworthy that, in a seemingly paradoxical way, the farther from the sea such narratives about the flood are recorded, the more pregnant is their sea symbolism (Poruciuc 2006).

The basic theme of all flood myths that have been written down stems from earlier narrative versions. The oral preservation of mythic stories is a collective process in which items of memory are typified. The memory of concrete events are also stereotyped in this way. What we find in the flood myths is not an accurate account of the Great Flood of ca. 6700 BCE, but a typified construct of the event that lives on in manifold variations. In recent years, folkloristic research has produced findings that testify to the role of variation as a major arbiter in all thematic issues of the oral and written traditions (Honko 2000).

Variation is relevant not only in an intercultural comparative framework of thematic settings in myths, but also in the concrete recording of an individual text (oral or written) in the horizon of time. It is practically impossible to reconstruct, from existing oral versions of a myth to a proto-version dating to prehistoric times. Therefore, we will never know with any certainty which were the first thematic and local variations of the flood myth that emerged after the Great Flood because we have to assume that variations were lost in the translation from one generation of storytellers to the next, even from one performance to the next. And yet myths originate from stories that assume a typified form.

Stories function as memes, that is, clusters of ideas that remain on people's minds. If we take stories as an example, a story that has great emotional impact, or for any other reason has the effect that you just cannot stop

thinking about it, will go round and round in your head. This will consolidate the memory for that story and will also mean that, since you are thinking about it a lot, you are more likely to pass it on to someone else, who may be similarly affected. (Blackmore 1999: 40*f*.)

The Black Sea catastrophe must have had a tremendous impact on people's cultural memory and it created a mythic tradition with a wide range. Eventually the theme of the Great Flood found its thematic representation in the canonic version of Noah's flood and, with the biblical tradition, it spread over the whole world. People in many countries around the world know the flood theme through their Christian or Judaic education, or they have heard about it through idea diffusion. The myth of the Great Flood is among the most successful themes in human history.

The variety of narratives about flood events that have been collected by scholars of mythology is perceived by some as indicative of the fanciful nature of flood stories and they dismiss any connection of the contents of a myth with geologic realities. It is true that a flood myth cannot tell you anything about when or how exactly a real event occurred and how long it lasted. On the other hand, the value of a myth for science is not limited to such negative features. In the core of a flood myth may be encapsulated the cultural memory of a real geologic event. In this perspective, a flood myth may be the culturally modulated manifestation of a distant echo of what really happened.

The dismissal of flood myths—of the biblical and Sumerian flood myths in particular—as lacking any real core on grounds of their wide occurrence in the cultures of the world is a pseudo-argument. Modern investigations of flood events that are inspired by pertinent allusions in local flood myths defy such dismissal. Recall that, in the prescientific world of archaic Greek society (eighth century BCE), the original meaning of myth, Greek *muthos*, was a mythopoetic account of what is true, that is distilled knowledge about matters of significance, handed down from one generation to the next.

Then, *muthos* was not associated with notions of the imaginary or with entertainment and it was not opposed to *logos* as the term referring to reason and logical thinking. In the context of Homeric epic literature, myths (*muthoi*) were synonyms of holy ideas (*hieroi logoi*). Thus, in archaic Greece, *muthos* was thought of as referring to the cultural knowledge that had been accumulated in society.

The connotations that *muthos* had in archaic Greek society were revitalized and formed part of the philosophical thought since the times of Plato (ca. 427–ca. 347 BCE) who restored the original meaning of *muthologia* as tradition, or the useful knowledge handed down from the ancestors. Mythology as a modern science investigates the embedding of the traditional knowledge that is transmitted in an ancient community through the medium of myth.

Where, When, and How Did the Black Sea Flood Occur?

For a very long time, the origins of the Black Sea were shrouded in mystery. During the past 100 years, the historical conditions of this large water basin have been studied by natural scientists who made sensational discoveries near the turn of the 21th century. According to these new findings, the most dramatic geologic events experienced in the Pontic region were two massive inundations that occurred at different times. These flood events are named after those who discovered them.

The first was an inundation that took place between ca. 15,000–13,000 BCE (Chepalyga's flood). The spill water of this early flood was directed from the east (Caspian depression) via the Manych-Kerch spillway (900 km long, 10–50 km wide) to the Black Sea depression in the west (Chepalyga 2007). This inundation is evidenced by the remainders of freshwater fauna in the sediments on the shallow shelf of the Black Sea. The second was the inundation of ca. 6700 BCE (Ryan's flood) in which the spillway was directed from the south (Mediterranean Sea) via the Dardanelles and the Marmara Sea north to the Black Sea (Ryan and Pitman 1998).

These massive inundations caused long-term repercussions in the ecology of the regions around the Black Sea (i.e., in the circum-Pontic area), and the resulting environmental changes had an impact on the human ecology of the populations that lived in and near the area where flooding occurred as well as in a larger geographic region. From the perspective of archaeology and cultural history, the Pontic area can be divided into a southern and a northern zone. The northern Pontic zone is the area stretching from the lower Dnieper basin eastward to the northeastern coast facing the Caucasus. This region includes the Caspian depression and borders on the basin of the Middle Volga. The southern Pontic zone refers to the cultural area along the southern coast of the Black Sea extending to the west as far as the lower Danube basin. The southern Pontic zone includes the Balkan peninsula and Greece on the European side as well as Anatolia on the Asian side. The eastern part of Anatolia borders on the Caucasus as its geographic outpost. The easternmost part of the southern Pontic zone opens to the historical landscape of Mesopotamia.

Within the wider region, there has been cultural interaction, since prehistoric times, between the West (Europe and western Anatolia) and the East (southern Caucasus region and Mesopotamia). In a simplified way one can say that Chepalyga's flood was an arbiter of the formation of Mesolithic cultures in the northern Pontic zone, while Ryan's flood contributed to an acceleration of the process of Neolithicization (that is, to the spread of agrarian lifeways) in the southern Pontic area.

The geologic discovery in the 1990s of the Black Sea flood—"the event that changed history" (Ryan and Pitman 1998)—has spurred a lively debate about the changing ecological conditions in the eastern Mediterranean region. These findings, which continue to accumulate, challenge traditional patterns of research on Old World civilizations (Haarmann 2002, 2006). How did the flooding of the

prehistoric freshwater Euxine Lake and the extension of its coastline to the present size affect the surrounding environment, and what was the human response to the ecological changes that occurred?

Ryan and Pitman (1998) initially dated the catastrophe to ca. 5600 BCE. In light of new geologic evidence, the date had to be revised to ca. 6700 BCE, and the new date was first announced by Ryan in a personal communication to me in June 2002. A crucial issue in connection with the flood hypothesis is the change in salinity of the waters connecting the Black Sea and the Mediterranean. The dating of the flood is intrinsically associated with the measurement of changes from a state of fresh water in the prehistoric Euxine Lake to the diagnostic value for global sea water in the newly emerged Black Sea (Ryan et al. 2003).

The new time frame corroborates the sensational insight that the process of Neolithization in the southern Pontic zone, especially in southeastern Europe, accelerated around the mid-seventh millennium BCE. A period of great ecological change began with the warming of the climate toward the end of the last Ice Age, eventually causing a sea-level rise from melting ice in the Mediterranean and, at the same time, a desiccation of the Pontic area. How can these processes, which stood in opposition to each other, be explained?

As a result of the melting of the continental glacier after 11,000 BCE, several large freshwater lakes emerged, most of them in the north. Lake Ladoga and Lake Onega are the remnants of these ancient lakes. In the south, the Caspian Sea, Lake Aral, and the precursor of the Black Sea (New Euxine Lake) were formed. The water reservoir in these lakes steadily increased with the flow of water melting from the ice shield. The prehistoric Danube, Dnieper, and the Don rivers provided most of the water for the great southern lake.

Around 9400 BCE, a second meltwater spike began. The waters that were then released from the melting ice shield, however, cascaded into the Caspian Sea, into Lake Aral, and accumulated in the northern great lakes, but did not reach the Euxine Lake. The waters of the second spike did not flow south because the level of the Earth's crust south of the former glacier lay higher than the land that had been pressed down under the mass of ice. The meltwater pooled in this depression then flowed west into the North Sea and east into the Caspian Sea and Lake Aral. As a result, the Euxine Lake steadily evaporated.

Toward the end of the 10th millennium BCE, the level of the Euxine Lake had dropped below the level of the external ocean. Global warming after the Ice Age was the decisive factor for the rise of sea level in the great oceans and, particularly, in the Mediterranean Sea. By ca. 6700 BCE, the difference between the level of the Euxine Lake and that of the Marmara Sea was about 70 m. At that time, the absolute level of the Euxine Lake was 85 m below the present day sea level, which itself lay about 15 m lower than it is today (William Ryan, personal communication). It was only a matter of time before the collected mass of ocean water would break through the Bosporus Strait.

The flooding of the Black Sea basin may have been triggered by a major earthquake combined with volcanic activity in the Aegean (such as the one that ripped the island of Santorini apart ca. 1625 BCE), or by a series of major earthquakes like the ones that shook northwestern Turkey in the summer and autumn of 1999. Recall that the passage that links the Mediterranean with the Black Sea (i.e., the Dardanelles and the Marmara Sea) is the area of the world with most seismic activity during the past 30 ky (i.e., 30,000 years).

Once the land bridge separating the ancient Sea of Marmara and the Euxine Lake had broken under the pressure of the high-level water masses in the south, marine water began to cascade into the freshwater reservoir below "traveling at a speed of up to fifty miles per hour through the narrowest constrictions. The water rushing into the Black Sea would have raised its surface half a foot per day (Ryan and Pitman 1998: 160). The cascading of marine water through the Bosporus Strait into the former Euxine Lake continued for years. As a result of the flooding, the surface of the Euxine Lake widened and its waters became interconnected with the tides and flows of the ocean waters farther south.

The evidence for the catastrophic event of the flood, for the extensive inundations, and for the birth of the Black Sea comes from different disciplines, from geology, marine biology, and oceanography. Underwater archaeology, however, has so far not produced unequivocal results.

1. On a modern maritime map one can easily identify the ancient lake and its shoreline (Map 1). The former Euxine Lake was contained within a deep freshwater pit with a depth of more than 2000 meters with steep underwater
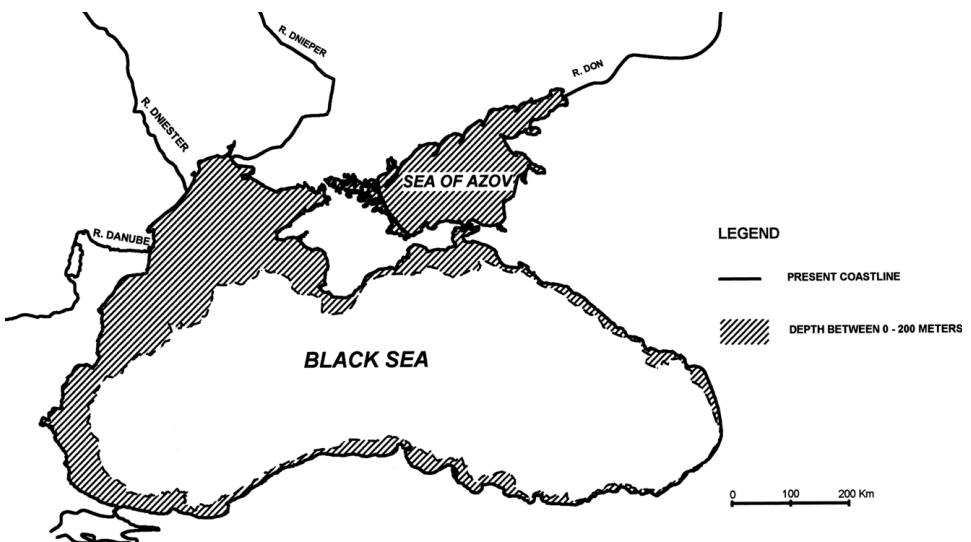
Maritime map of the Black Sea. (Courtesy Harald Haarmann)

slopes. This deep central pit is now surrounded by the post-deluge extension of shallow marine waters.

2. The shape of the marine passage near the Bosporus Strait is that of a ravine. In coastal waters, such a geologic profile could have only originated as the result of a cascading of inflowing water under high pressure.

3. Evidence for the direction of the flood (i.e., through the Dardanelles and the Marmara Sea northward) has been discovered in the sediments near the Bosporus Strait where the remnants of marine shells are strewn along the ancient trail of the flood waters.

4. Among the evidence for the thrust of the sea water that covered the former lake are the remainders of freshwater fauna in Black Sea sediments near the Bosporus Straits that were buried under the massive influx of sea water.

5. Geologists have discovered the ancient canyon of the River Don from the predeluge era as the result of electronic sonar monitoring of the sea bed (Ryan and Pitman 1998: 125). Before the flood, the ancient Don crossed an alluvial floodplain (now the Sea of Azov), crossed the Kerch Strait, and formed a delta on the flat shore of the former Euxine Lake.

6. Certain underwater phenomena provide further information about the nature of the ancient shoreline. Methane gas continues to be released from the remains of decaying plants. This residue of ancient land vegetation is indicative of the lagoons and swamps that separated the open lake from the low and fertile marshes of the northern plain (a terrain that is nowadays politically divided between Ukraine and Russia).

7. Proof for the catastrophic nature of the flood event comes from the quality of the marine water in the Black Sea. The former Euxine Lake was a freshwater reservoir that was literally buried under the inflowing masses of salt water. Where fresh water and salt water are in constant contact (e.g., the offshore waters of the deltas of big rivers that flow into the ocean), they gradually mix. No such mixing took place when the Black Sea originated. The salt water of the flood covered the freshwater reservoir of the Euxine Lake rapidly and cut it off from supply with fresh oxygen, resulting in the extinction of the fauna and flora in the ancient lake. The water in the deep central pit of the Black Sea is highly toxic.

The Black Sea is the world's biggest single reservoir of hydrogen sulphide. Below a fluctuating depth of between 150 and 200 metres, there is no life. The water is anoxic, without dissolved oxygen, and impregnated with H_2S; because much of the Black Sea is deep, this means that some 90 percent of the Sea's volume is sterile. (Ascherson 1996: 4f.)

Sometimes when storms disturb the upper level of the marine waters, it may happen that water from deeper levels spills to the surface. When this water

comes in contact with the hulk of a vessel, the surface of it (whether wood or metal) turns black. This phenomenon is the source from which this sea got its name. It was called Mare Nigrum (Latin for Black Sea) by the Romans and, later, the people living around the Black Sea continued this name-giving tradition (e.g., Russian Chernoe More, Turkish Kara Deniz).

A detailed picture of the loss of land caused by the Great Flood does not provide an estimate of the size of population that may have occupied the fertile coastal area of the Euxine Lake in the predeluge era. Underwater archaeology may one day discover some of the settlements that had to be abandoned ca. 6700 BCE. The search for remainders of predeluge settlements and human remains has been ongoing for several years. Looking for such remains is like trying to find a needle in a haystack. In many areas, underwater currents have reworked the profile of the marine landscape so that traces of human settlements or artifacts may be buried deep in the sand and mud, hard to discover.

For the time being, not even the most modern technological equipment has brought to light any solid evidence for predeluge artifacts or settlements. Perhaps the oceanographer with the greatest reputation for lucky discoveries is Robert Ballard who has enjoyed publicity ever since his discovery of the *Titanic* in the 1980s. Ballard carried out underwater investigations in the shallow waters of the Black Sea in September 2000. Off the shore of Turkey his team identified remnants of stone buildings and wooden structures. The dating of the wooden pieces, however, brought disappointment because the radiocarbon date did not match the early time frame of Ryan's flood. The underwater site that Ballard discovered had been flooded as the result of sea level change at a later period.

Numerous conferences have been organized to bring experts from different fields together to discuss crucial issues relating to the Black Sea flood. The biggest enterprise so far is the United Nations Educational, Scientific, and Cultural Organization (UNESCO) –sponsored International Geoscience Programme "Black Sea-Mediterranean corridor during the last 30 ky: Sea level change and human adaptation (2005–2009)," with experts convening at annual conferences. These conferences provide a forum for exchanging views on the pros and cons. There is consensus about the basic fact that, at a certain point in prehistory, marine waters from the Mediterranean spilled over and filled a former freshwater lake, thus linking the Black Sea to the Mediterranean. And yet, many details of this general scenario continue to be debated.

Human Response to the Black Sea Flood and How Cultural Evolution in the Postdeluge Era Was Affected

The flood had an impact on the climatic situation. A greatly expanded Black Sea water body had more surface area for evaporation and temperature modulation that led eventually to climatic changes, which, in turn, directly influenced

vegetation. Colder periods produced boreal forests, and in such a landscape the spread of farming was moderate. As the climate grew warmer after 5800 BCE, grasslands became more prevalent, which facilitated the spread of farming. The sudden increase in the number of new settlements around 5500 BCE resulted from environmental changes caused by the rapid warming of the climate.

Agriculture, the technology of plant cultivation, and agrarian lifeways had been adopted by the indigenous inhabitants of southeastern Europe already ca. 7000 BCE, that is before the occurrence of the flood. The turmoil caused by the large-scale inundations of arable land, especially in the northwestern and northern part of the Black Sea, must have triggered displacement of sedentary communities and local migration movements. How did humans respond to the flood event and its after-effects? How did they reorganize their community life, and what effects did the flood have on cultural evolution? These problems have been addressed by scholars from different scientific disciplines of the humanities who gathered at the first conference on the interdisciplinary significance of the Black Sea flood, sponsored by the Bogliasco Foundation and the Institute of Archaeomythology (Sebastopol, USA), which was held in Bogliasco (Italy) in 2002.

When investigating the after-effects of the Black Sea flood within a wider perspective, one has to be aware that the flood is but one event among others that changed the ecology in the circum-Pontic area. The decisive effects on the geology and the climate of the Pontic region stem from a series of events that changed the ecological balance of the region: the Great Flood of ca. 6700 BCE; a mini Ice Age that lasted from ca. 6200 until ca. 5800 BCE; and a radical warming which began ca. 5800 BCE.

This sequence of events, dated by absolute chronology, provides a temporal frame for a reconstruction of human response to radical climatic transformations that had manifold consequences, ranging from the reshuffling of local economies to the improvement of various technologies. In its totality, this sequence of events and their repercussions in human ecology constitute something like a "big bang" in cultural evolution, induced by environmental changes, which, in turn, affected changes in human ecology:

- ca. 7000 BCE Settlements with agrarian population in Anatolia (Çatalhöyük), in Southeastern Europe (Knossos in Crete; Nea Nikomedeia in Macedonia)
  —Settlements on the shores of the former Euxine Lake

- ca. 6700 BCE The Great Flood (the Black Sea catastrophe); flooding of the former Euxine Lake

- ca. 6500 BCE New settlements emerge in Southeastern Europe (Sesklo, Achilleion, Starcevo)
  —Increase of cultural activity in the area

- ca. 6200 BCE Beginning of a mini Ice Age
  —Stagnation of regional settlements

- ca. 6000–5500 BCE Dry weather in Anatolia; arable land dries out
  —Settlements such as Çatalhöyük and Hacilar are abandoned
- ca. 5800 BCE Beginning of a radical warming of the climate
  —Rapid expansion of agrarian settlements in southeastern Europe
- ca. 5500–5000 BCE The early stage of civilization in southeastern Europe (the Danube civilization as a trade-oriented commonwealth, with advanced technologies and institutions)
  —The beginnings of sacral architecture, metal-working, writing; technical specialization of crafts such as pottery-making and weaving; proliferation of artistic expression (e.g., production of figurines in different types and styles)

In the course of the seventh millennium BCE, human ecology was characterized by changes that suggest a marked distinction between the predeluge and postdeluge eras. Commenting on developments in southeastern Europe, Bailey writes: "What is common across the regions, however, is that the changes mark off the post-6500 BCE period from the preceding millennia" (2000: 39). It is noteworthy that, also in the northern Pontic region (in areas north of the Black and Azov seas), the mid-seventh millennium BCE marks a fundamental change, the transition from the Mesolithic to the Neolithic period (Telegin 2002: 37ff.). Cultural life in Europe drifted away from that in Anatolia because of its own rapid pace. The areal sociodemography of settlements and the role of the mid-seventh to mid-sixth millennia BCE were turning points in cultural chronology. What developed during this period can be considered civilization in terms of high culture.

In Anatolia, the climatic changes of the seventh millennium BCE (the mini Ice Age) and of the sixth millennium BCE (the rapid warming) lead to the abandonment of major settlements, while, in southeastern Europe, ecological conditions favored the extension of the area of settlement. The earliest urban settlement in Anatolia, Çatalhöyük, inhabited by an agrarian population, was abandoned ca. 6000 BCE. This was the time when, in the western part of the southern Pontic zone, the settlement of Karanovo (east central Bulgaria) was founded. This site would become a major cultural center and is name-giving to one of the main regional cultures within the Danube civilization. Around 5500 BCE, Hacilar in western Anatolia was deserted by its inhabitants, possibly due to increasing aridity of the area, which made farming exceedingly difficult. On the European side of the Pontic zone, the settlement of Vinca near Belgrade (Serbia) was founded on the southern bank of the Danube at about the same time (ca. 5500 BCE). Vinca would become the most important center of trade relations and cultural development in the central Balkans, and this site is name-giving for the western major regional culture within the Danube civilization (Gimbutas 1991: 62ff.).

References and Further Reading

Ascherson, Neal. *Black Sea. The Birthplace of Civilisation and Barbarism*. London: Vintage, 1996.

Bailey, Douglass W. *Balkan Prehistory: Exclusion, Incorporation and Identity*. New York: Routledge, 2000.

Chepalyga, Andrey. "Epoch of Extreme Inundation in the Ponto-Caspian basin as Noah's Flood new concept." presented at XVII International Union for Quaternary Research Congress, 2007.

Blackmore, Susan. *The Meme Machine*. New York: Oxford University Press, 1999.

Boyle, Katie, Colin Renfrew, and Marsha Levine, eds. *Ancient Interactions: East and West in Eurasia*. Cambridge: McDonald Institute for Archaeological Research, 2002.

Dalley, Stephanie. *Myths from Mesopotamia. Creation, the Flood, Gilgamesh, and Others*. New York: Oxford University Press, 1998.

Gimbutas, Marija. *The Civilization of the Goddess: The world of Old Europe*. San Francisco: HarperSanFrancisco, 1991.

Haarmann, Harald. "On the Formation Process of Old World Civilizations and the Catastrophe that Triggered It." *European Journal for Semiotic Studies* 14 (2002): 519–93.

Haarmann, Harald. "On the Fabric of Old World Civilizations: Human Response to the Black Sea Flood and Subsequent Climatic Change." *Journal of Archaeomythology* 2 (2006): 27–64.

Haarmann, Harald, and Joan Marler. *Introducing the Mythological Crescent. Ancient Beliefs and Imagery Connecting Eurasia with Anatolia*. Wiesbaden: Harrassowitz, 2008.

Honko, Lauri, ed. *Thick Corpus, Organic Variation and Textuality in Oral Tradition*. Helsinki: Finnish Literature Society, 2000.

Keay, John. *India—A History*. London: Harper Perennial, 2000.

Marler, Joan, and Harald Haarmann, eds. "The Black Sea Flood: An Interdisciplinary Investigation." *Journal of Archaeomythology* 2 (special issue), 2006.

Oppenheimer, Stephen. *Eden in the East: The Drowned Continent of Southeast Asia*. London: Phoenix, 1998.

Poruciuc, Adrian. "The Sea and the Sea-Flood Motif in Romanian Folklore." *Journal of Archaeomythology* 2 (2006): 65–76.

Ryan, William, and Walter Pitman. *Noah's Flood. The New Scientific Discoveries about the Event that Changed History*. New York: Simon and Schuster, 1998.

Ryan, W. B. F., C. O. Major, G. Lericolais, and S. L. Goldstein. "Catastrophic Flooding of the Black Sea." *Annual Review Earth and Planetary Sciences* 31 (2003): 525–54.

Telegin, D., et al. "The Chronology of the Mariupol-type Cemeteries of Ukraine Re-visited." *Antiquity* 76 (2002), 356–363.

Yanko-Hombach, Valentina, ed. *Black Sea-Mediterranean Corridor during the Last 30 ky: Sea Level Change and Human Adaptation* (2005–2009). UNESCO–IGCP–IUGC–The International Geoscience Programme, Extended Abstracts (October 2005). Istanbul: Technological Research Council of Turkey, 2005.

Yanko-Hombach, Valentina, et al., eds. *The Black Sea Flood Question. Changes in Coastline, Climate and Human Settlement*. Heidelberg: Springer, 2006.

CON

Throughout the world, the story of Noah and the Great Deluge has reappeared and spread across the world in various cultures and myths. This story of a man, who responds to the warning of a universal flood, builds a rescue vehicle, and survives the flood, repeats itself in many cultures and civilizations. This myth raises many questions concerning the possibility of putting many animals in one ark, the cleaning of the ark, and feeding of the animals on the ark. Some theorists believe that the myths are remnants of a vast universal flood that has covered the world at a certain time. It might have been a worldwide tsunami that flooded across the world. They have offered fossil evidence of a widespread flood that covered large territories of land with water. Others have assumed that there might be archaeological evidence of the ark's existence on Mount Ararat. Other theorists like Joseph Campbell believe the universal flood to be representative of psychological themes that surface from the subconscious mind in the stories of mankind. There are other Noah theorists who discount the improbability of the flood and the building of the ark. They posit that it would be impossible to house all these animals and plants in the ark, challenge that the ark would be seaworthy, and question whether Noah could have been too old, because it took 120 years to build the ark. It is these themes that we will explore in this section of the chapter and the psychological interpretations of them representing the deep urges of man's psyche to survive and integrate disasters in primitive people's thoughts and ideas of the world.

The Western version of the myth has numerous forms where a righteous man, Noah, preaches against the sins of the world and warns of a future universal flood that will wipe out mankind. He preaches for 120 years to give mankind time to repent of their crimes. Noah exclaims of God's wrath against the sexual sins, crime, and inhumanity of the human race. Meanwhile, Noah persuades his family to build an ark to house the animals and his righteous family.
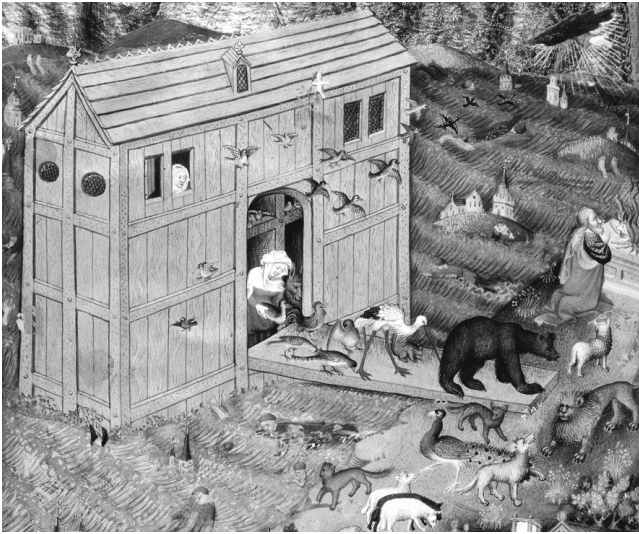
A depiction of the animals leaving Noah's Ark after the Great Flood, from the *Bedford Book of Hours*, about 1423. (British Library/StockphotoPro)

When the flood arrives, the ark survives through the storm for 40 days and nights. Then Noah releases a dove to discover dry land at least three times and he is finally relieved to find an olive branch in the dove's mouth. Noah sacrifices to his God a lamb for allowing him to survive the flood. The story of Noah further goes into Noah getting drunk and being discovered naked by Ham. Noah is regarded as the discoverer of wine and wine drinking by the Jews.

In the eyes of religion, Judaism regarded Noah as a righteous prophet and through its traditions believe that he was born spiritual with white hair on his body. The Catholic Church interpreted the myth to mean that the ark is the church that is escaping the flood that represents the world of sin. Saint Augustine of Hippo (500 CE) symbolized the church in the Noah's flood story. He actually measured the dimensions of the body and compared it to the ark to see that it represented the body of the church. The Mormons believe Noah to be the angel Gabriel, who would announce repentance to the world. These are the various religious interpretation of the original myth. Many churches who believe in the literal interpretation of the myth have attempted to prove it as an archaeological fact.

Others viewed myths as actual reflections of natural events or occurrences. The school of catastrophism believed that myths reflected natural disasters and floods. Immanuel Velikovsky (1895–1975), psychologist and explorer, wrote in *Worlds in Collision* that Earth suffered impact from a Saturn ejection of planetary matter onto Earth when there was a flood. He believes that myths are oral histories that reflected cosmic events such as a star explosion or a planetary explosion in outer space. Earth was influenced by electromagnetic events in outer space (Velikovsky, 1950).

Historical Proofs of the Debate

Many psychologists and anthropologists have attempted to pose the issue of why the myth has so many widespread versions around the world. This myth

has been discovered in various other cultures and belies the originality of the biblical version of the myth. According to some authors, there are as many as 600 stories of the flood around the world. They may prove the dispersion theory that the flood myth originated in one place and diffused around the world to various cultures in the world. The other is the local or regional theories that state the flood originated locally especially around the Black Sea. The Old Testament story has originated in various other mythological stories in earlier versions. The original Sumerian myth (third millennium) of Ziusudra tells how the god Enki warns Ziusudra that the gods will destroy mankind and Enki tells Ziusudra to build a large boat that allows him to survive for seven days. Ziusudra makes sacrifices to Enki and Ann and is given eternal life by them. In the Mesopotamian myth of *The Epic of Gilgamesh*, Utnapshtim builds a vessel to escape the vengeful Enlil, who sends a flood to destroy mankind. In the Babylonian myth, the gods destroyed the world three times with plague, famine, and then a flood. Enki received a warning from Ea and built a boat to escape the flood. The Akkadian myths are found in the Atrahasis Epic (1700 BCE) where Atrahasis is warned of a flood, because of human overpopulation. Enlil is disturbed by the noise of the humans and he sends the floods. Every 1,200 years, the gods send the flood, but Enki tells Atrahasis, who escapes the flood. Enki proposes to the gods to cause mankind to have miscarriages, to increase infant mortality, and to allow barrenness as solutions for overpopulation.

In Europe, the Greeks, Romans, Celtics, Lithuanians, and Volga had their myths. Even the neighbors of the Hebrews had flood myths. The Chaldeans had the 10th king, Xisuthrus, who was ordered by Chronos to build a vessel and sail during the flood. He sent a bird out twice and then he ran his ship aground on the Corycean Mountains (Frazer 1919: 108–10; Smith 1873: 42–43). Then in Greco-Roman mythology (1529 BCE), Deucalion and Pyrrha escape in a box-shaped boat to escape a flood. The Celts had a myth about heaven and earth, who were giants that crowded humans between them. One of the sons of man led his brothers to cut heaven into many pieces and spilled blood that created the floods. One of the Titans saved a pair of humans by building a ship (Gaster 1969: 93).

In Central America, the Tarascan, Yaqui, Totanac, and Maya have several myths. The Puzob dwarf people were destroyed by god because they neglected their offerings. They put some stones in a pond that saved some people. Four angels came down and became doves. Some angels ate dead animals and became buzzards (Dundes 1988: 194).

In South America, New Guinea, Venezuela, Colombia, Ecuador, Peru, Chile, tip of Argentina and Brazil as well as the Aztecs have many myths of the flood. The Aztec Calendar Stone describes a flood of 52 years after four stages of mankind civilization. In the nation of Ecuador, two brothers escaped a flood and went down the mountain Huaca ynana. In order to survive, they trapped two birds in a cave, which had cooked food for them in their hut. They captured

them and one of them married the brother to perpetuate the species (Frazer 1919: 268–69).

> Then by the will of Hurakan, the Heart of Heaven, the waters were swollen and a great flood came upon the manikins of wood. They were drowned and a thick resin fell from heaven. Because they had not thought on Hurakan, therefore the face of the earth grew dark, and a pouring rain commenced, raining by day and by night. Then ran the manikins hither and thither in despair. They climbed to the roofs of the houses, but the houses crumbled under their feet; they tried to mount to accomplish the ruin of this race, destined to be overthrown. (Spence, 1908)

In the continent of Asia, China, Korea, Mongolia, India, Burma, Philippines, New Guinea, and Malay Peninsula all have flood myths. Then in the Chinese *Book of Documents*, the Emperor Yu (2250 BCE), one of the prehistoric emperors, battles flood waters to save mankind from the devastation. The Indian myth of Vishnu's incarnation as the Matysa fish, who rescues Manu from a worldwide flood, repeats the same theme. Across the world, there are various myths with the same motif of a general flood of waters and the last man to escape the devastation of this problem. In Thailand, there was a mouse that started a flood and the human race was warned by other gods. They built a ship that survived the flood and they founded the human race.

In the Polynesian culture, the natives of Papua New Guinea believe that a worldwide deluge will rise, with Mount Teluga being the haven of safety. The Polynesian islands have several myths of the flood. In Tahiti, the sea god was angered by fishermen's hooks being entangled in his hair and he sent a flood to wipe them out. The Samoan fire and water gods flooded the world in their war between themselves.

In the American continent, the Indian tribes have many different types of myths. In the United States many tribes have myths such as the Netsilik Eskimo, Tlingit, Hareskin, Tinneh, Haida, Squamish, Skagit, Skokomish, Shasta, Nisqually, Warm Springs (Oregon), Joshua, Northern California Coast, Salinan, Luiseno (Southern California), Yakima (Washington), and the Spokana. The Blackfoot Indians of North America recount an Old Man in boats of water, who sends three animals that dig mud from the bottom of the ocean to rebuild the earth. The Navajo people had their gods expel the Insect people from a lower world through a wall of water. The Insect people flew up to the second world. Then in the fourth world, the Insect people's children were likewise punished and they went to the fifth world where people live today. In the Hopi tradition, the human race alienated them from Sotuknang, who had tried to destroy the world by fire and by cold. The world became evil again and so Sotuknang flooded the world, but helped several groups build a reed ship to escape the

deluge. They were guided by receiving wisdom from the top of their heads to find other places of habitation.

Australia has several myths of the great flood. Grumuduk, a medicine man, brought rain down from power of medicine. A plains tribe kidnapped him to get his powers, but he escaped and said water would rise in his footsteps that flooded on the enemies (Flood 1983: 17).

In opposition to the archaeological side of the debate, many believe that the myths of flooding are common to many regions and result from a deep-seated common human psychological source. This theory believes that the flood occurred at various times and in various regions, which shows that a universal flood could not have happened at one time. These theorists believe that the myth of the flood originated from the instinctual and subconscious needs of the minds of the mythmakers. They believe that when the shaman goes into a trance, a prophet sees a vision, or priests have inspirations, their source of beliefs and stories arises from their subconscious needs and motivations that are express in these psychic experiences. The needs and motivations are expressed in various symbols, designs, and patterns in their religions and superstitions. As a result, the myths of mankind originate from the subconscious needs of mankind as expressed in various patterns in the stories of the myths as well as the symbols used in the myths. Some of them believe that we all share a universal subconscious that derives from the racial memories of our ancestors and they are contained in the minds of men. These memories may take the form of dreams, visions, and inspiration that are common to all stories, literature, and poems. When men explored the various cultures of the world, they uncovered various similarities between cultures that could not be explained as coincidences, so they sought to discover the origins of these similarities in the common cultural consciousness of mankind. We all share similar cultural needs and challenges, so these have imprinted themselves on men's minds. As a result, these cultures have symbolized and written them in their myths and literature.

Historical Interpretations

Since the early 18th, and into the 19th, century, mythology and comparative mythology became important fields of studies. Many scholars noted similarities of myths and gods across the human spectrum and they attempted to find patterns. These schools posited distribution patterns that show how myths could be spread across the world. The first pattern was one of diffusion from a central location like Mesopotamia and the other was one of local spots of diffusion that spread to several areas. The myth and ritual school became established during the 1800s, representing myth as a part of the ritual system that united societies. This school consisted of people such as William Robinson Smith, James Frazier, Jane Ellen Harrison, and S. H. Hooke. It began with the 1894 publication of James

Frazier's *The Golden Bough*, which analyzed myths to find various functions in outlining taboos and cultural norms for the primitive people. Myths served various needs of the people from justifying their various institutions to justifying their history, religions, and customs with heroes such as Noah (Frazier, 1894). Frazier's theory is that man attempted to understand natural law and utilized magic to understand or control these natural laws. Because of the growth of magic, religion and myth embodied man's search for magic and spells to affect other people and the universe. The study of myth exploded after this with the publications of books comparing religions and myths throughout the world.

In the 1930s, the Soviet school developed, which derived myth as representing the traditions of folklore and worldview of the culture. Recently, there is a dispute on the issue of whether myth was primary or whether there were parallel paradigms between myth and ritual.

There also grew a psychoanalytic school that proposed that myths originated from the human subconscious mind due to fears and anxieties of the human bodily system. These myths were supposed to teach people about how to live and escape danger. Carl Jung (1875–1961), German psychologist, theorized in his studies that men had a universal subconscious that represented the instinctual reaction to the world. He believed that all men had these themes, motifs, or symbols that repeated itself in the cultures of the world. They could be seen in the dreams and nightmares of men as well as the spiritual traditions

---

An Alternative Scientific Theory on the Deluge

One scientific alternative to the idea of the Black Sea deluge has its basis in an associated theory called "crustal displacement," which was proposed in 1958 by U.S. scholar Charles Hapgood, whose ideas on the causes of early historical geologic cataclysms (such as great floods and volcanic activity) gained the support of Albert Einstein, who wrote the foreword to Hapgood's book *Earth's Shifting Crust: A Key to Some Basic Problems of Earth Science*. Essentially, crustal displacement theory holds that the Earth's crust moves rapidly in one piece, thus causing a large number of cataclysmic events to occur over a very short period of time. According to Hapgood, such a crust shift occurred around 9600 BCE, during which he suggests that a large part of the lithosphere that supports the tectonic plates shifted violently over a period of just days or weeks. Such shifts, Hapgood argued, cause the locations of various features to move dramatically. This type of phenomenon would inevitably cause a large displacement of seawater, leading to a great deluge that ancient peoples might consider to have come from on high.

Sources: Hapgood, Charles H. *Earth's Shifting Crust: A Key to Some Basic Problems of Earth Science*. New York: Pantheon Books, 1958; Hancock, Graham. *Fingerprints of the Gods*. New York: Crown Publishers, 1995; Flem-Ath, Rand, and Rose Flem-Ath. *When the Sky Fell: In Search of Atlantis*. New York: St. Martin's Press, 1995.

---

of the world. Carl Jung believed that these archetypes symbolized the conflicts, instincts, and responses to the energies of the body. He also believed that through analysis you could give meaning to a person's psychological life and therefore balance his or her conflicts as well as psychological disturbances.

In Jung's (1964) famous book *Man and His Symbols*, he discusses the historical analysis of various symbols in the myths of mankind. He cites various archetypes such as the nurturing Mother, the Hero, the Trickster, and other tales of the subconscious mind. Jung shows that his psychiatric patients drew pictures that represented these archetypes in their representations of their conflicts and desires. These representations have also been found in the Noah myth as being symbolic of people overcoming their dark urges from the subconscious. This he believes is a reflection of the psyche's need to express itself into the subconscious.

Mircae Eliade (1907–1986), famous comparative religionist, has also attempted to compare myths and religions as expression of man's need for sacredness and peacefulness. He has attempted to show that men have tried to find sacred space, sacred time, and sacred resolution of trials. Eliade interprets the Noah myth in the context of Christian symbolism as representing the meaning of water as formlessness and thus represents creation and birth. On the other hand, it also represents the loss of form or rebirth, a state prior to birth. The man is returned to his origins in water. So when Christianity absorbed this as a baptism motif, Noah became symbolic of mankind's baptism and return to a sacred situation before his creation. In Eliade's (1963) *Myth and Reality*, he explores how Noah has been symbolic of man's search for sacredness amidst trials and tribulations.

One of the most famous analyzer of the psychological roots of the Noah myth is Joseph Campbell (1904–1987). Campbell, a comparative mythologist, has written several books on comparative mythology utilizing Jungian analysis, comparative mythology techniques, and ritual processes to analyze myths around the world. Throughout his analysis, Campbell believes that there is a monomyth, a general mythical pattern, which applies to all the religions and mythologies in the world. In his explorations through myth, a central theme is reiterated through his writings, which are that myth is the reflection of the subconscious in dreams, visions, and whisperings of the spirit of man. He believes they are structured spiritual messages that reflect the energies of the psyche in the mind. They are messages that become symbols or archetypes in the hidden mind of the shaman, the spiritual leader, or the saint. Myths are representations of the universe itself in its various powers as described by religions as Melanese *mana*, Hindu as *shakti*, Chinese as the *chi*, and Christians as the power of God. Thus, the myths are representative of the Jungian idea of a universal subconscious that is shared by all mankind. This is a Jungian concept that hypothesizes that we all share a universal energy that manifests itself in unconscious urges and through archetypes in dreams of various patterns (i.e., mother, axis mundi, and hero myth). As

a result, all myths and religious stories are just representatives of these innate ideas that come from a universal source.

In the description of myth, Campbell shows the psychological interpretation of myths as symbols of the unconscious energy in our bodies as expressed in the forms of dreams and visions. We as embodied expressions of the universe's energy encapsulate these energies into the dream archetypes in our subconscious minds. These dream archetypes are common patterns, symbols, or motifs that repeat themselves in the myths and religions of the world as well as in individual dreams. Jung hypothesized that we all have a universal subconscious that we share as beings of nature, so we express the same cultural archetypes. As a result, our common subconscious energies are the same and have expressed themselves in the cultural structures of our society, namely in mythological storytelling, religious scriptures, and poetic forms that embody our dreams and inspiration. The muses, goddesses of art and music, of the creative life derive from the bodily energies in our subconscious that cry for expression and embodiment in the artistic forms of life. Finally, these bodily energies inform the cultural values and wisdom of every civilization on Earth through these literary and artistic forms created from the inspiration of the subconscious mind.

In his *Hero with a Thousand Faces*, Campbell analyzes the archetype of the hero in the world's religions and applies this archetype to the myth of Noah and world dissolution (1989: 50). In the archetype of the hero, the hero starts his journey by initiation into the call, trials, meeting with the goddess, reconciliation with the father, return to the world, refusal to return, and the bestowal of the boon. In the initiation stage, Campbell sees that hero is called to his quest by unrest in his society or the universe. He begins his quest, but some hero refuses the quest's call only to be called to a greater purpose by the universe. Then the hero finds a boon or gift of knowledge where he gains enlightenment about the universe. His self-consciousness is destroyed by the insight of the universal consciousness or his duty to a higher purpose. The hero is also tested by a series of intense trials and obstacles before he receives the boon. The boon is a gift of knowledge or spiritual insight. After receiving this enlightenment or boon of knowledge of the archetypes, the hero is then called to return to the real world and confronts the reentry into the world of forms. Afterward, he returns to teach or give that boon back after many trials and challenges. His return is met with skepticism until he gains students, who recognize archetypes in him. In this journey, Noah represents a negative hero who is called to face the trials of the deluge and who survives to bring life again to a cleansed world after being washed by the flood.

In his explorations in 1989, and other writings, Joseph Campbell criticizes the actual fact that the flood was a universal event. Joseph Campbell takes issue on the hypothesis that the flood was a geologic event. He states that the record of the flood in some city–states contain evidence of floods occurring at different

time periods. Campbell explains that those floods of Shuruppak and Uruk were written during the Shamir-Jasmiz period, 3000 BCE, the ones at Ur occurred during the Obed period, a millennium before, and that Kish occurred a third generation before. Thus, Campbell interprets these floods not to support a general flood hypothesis occurring at one time, because it would be challenged by the occurrences at these different times.

Campbell also discovers three types of myths about the Great Flood. The first type of mythological type is the planting of culture in different ages—ages of gold, silver, bronze, iron, where the moral condition is destroyed. The flood eradicates the old and brings a new start. The second mythology type is that God created the human race, some who misbehave. This God evolves from one who creates to one who becomes involved in the lives of the people. This God regrets what he has done in creating such a sinful race, and then he destroys them in a great flood. The mythic imagination shows the gulf between a previous prehistoric past and its transition into the present age (Campbell 2001: 43–44).

Joseph Campbell also states that it might have been a psychological or spiritual symbolism that involved the hero being overwhelmed or the city–state being overwhelmed by chaos and then being renewed in the flood of emotions and loss of balance. Campbell explains that there was no universal flood, but he challenges us that there is a spiritual interpretation of the story. If we attempt to read them as facts, they become dry and stale. What is the significance of the flood? Campbell says that it is the arrival of chaotic circumstances, the imbalance of nature, or the final curtain of an age. It is an end of a psychological stance or position that changes the person's character (Campbell 1990: 67–68).

Campbell applies psychological terms to the meaning of the monomyth. He shows that the universal flood symbolizes a disturbance of normal society, world existence, or even world meaning. Campbell describes the flood as representing the psyche, which is overwhelmed by lust, anger, or some emotional imbalance and then obtains control or management of these unconscious urges. Then the chaos or confusion occurs in the myth that represents conflicts in the person's life. In addition, he states that it is an end of a stage of life in the person. Finally, it renews itself into another expansion of consciousness. The subconscious is integrated with the superconscious by the myths of the hero and Noah.

Then Campbell offers another interpretation of the monomyth that the deluge motif is the annihilation of the ego system and that it is overwhelmed by his own nature. He explains that this flood pattern represents the destruction of the ego structure in the human psyche. The hero's conscious habits and realities are being destroyed. His human nature overwhelms the limits that are imposed by the disciplined society. His self is overwhelmed by the primitive urges of his subconscious mind, but he emerges as a hero who has conquered the limitations of his ego structures, and he becomes conscious (Campbell 1989: 36). The hero represents the overcoming of obstacles and challenges.

In conclusion, Campbell states that the myth of the flood represents the overcoming of obstacles either by the hero or by the society itself. It might have originated from local floods, but it was taken as a symbol of chaos and insecurity as well as the indomitable survival of the human species. The type of hero represents the renewal of human society and its survival from challenges and problems. Therefore, the Noah myth represents the subconscious energies of the human soul that pervade the universal religious spirituality of man and not an actual archaeological fact. These myths have affected the development of literature, movies, and fantasy stories in the modern era. It has also been used to analyze and help patients in the development of depth psychology and transcendental psychotherapies that utilize dreams and symbols to interpret their psychic depths.

References and Further Reading

Campbell, Joseph. *The Hero with a Thousand Faces*. Chicago: University of Chicago Press, 1989.

Campbell, Joseph. *An Open Life*. New York: Harper Perennial, 1990.

Campbell, Joseph. *Thou Art That: Transforming Religious Metaphor*. New York: New World Library, 2001.

Dundes, Alan, ed. *The Flood Myth*. Berkeley: University of California Press, 1988.

Eliade, Mircea. *Myth and Reality*. New York: Harper and Row, 1963.

Flood, Josephine. *Archaeology of the Dreamtime*. Honolulu: University of Hawaii Press, 1983.

Frazer, Sir James G. *Folk-Lore in the Old Testament*, Vol. 1. London: Macmillan, 1919.

Frazer, Sir James G. *The Golden Bough: A Study in Magic and Religion*. New York: Macmillan and Co., 1894.

Gaster, Theodor H. *Myth, Legend, and Custom in the Old Testament*. New York: Harper and Row, 1969.

Jung, Carl. *Man and His Symbols*. Garden City, NY: Doubleday, 1964.

Smith, George. "The Chaldean Account of the Deluge." *Transactions of the Society of Biblical Archaeology* 2 (1873): 213–34.

Spence, Lewis, ed. *The Popol Vuh: The Mythic and Heroic Sagas of the Kichés of Central America*. New York: AMS Press, 1908.

Velikovsky, Immanuel. *Worlds in Collision*. London: Gollancz, 1950.

4

Much of what is now considered to be Classic culture actually has Afroasiatic roots.

PRO   Talaat Shehata
CON   Harald Haarmann

PRO

Creating a firestorm in academic circles in the late 1980s and early 1990s, Martin Bernal's first two volumes of *Black Athena* questioned the authenticity and motives of European and American archaeological and academic research. The first volume, published in 1987, was titled, *Black Athena: Afroasiatic Roots of Classical Civilization: The Fabrication of Ancient Greece 1785–1985*. The second volume was published in 1991 and was titled *Black Athena: Afroasiatic Roots of Classical Civilization: The Archaeological and Documentary Evidence*. In 2006, a third volume was published to apply linguistic measures to support the arguments put forth by the first two volumes. The volume was titled *Black Athena: The Afroasiatic Roots of Classical Civilization*: *The Linguistic Evidence*.

Instead of using the usual references and attributing his view to an Afrocentric worldview versus a Eurocentric worldview, the informed reader could decide that the wisest path one needs to pursue during the 21st century and beyond is to adopt a world-centrist perspective. The tribal factors that often have a way of bleeding into such profound issues retard the intellectual development of the participants to the discourse instead of allowing for a constructive process of empathy and, ultimately, in some cases, enlightenment. Simply stated, Martin Bernal, an academician who just happens to be white, through rigorous research, found that his conclusions support the evidence that the well-known black African intellectual Cheik Anta Diop promulgated in the early 1960s, in his equally highly debated books *The African Origin of Civilization: Myth or Reality* (1974) and *Civilization or Barbarism: An Authentic Anthropology* (1991). In them, Diop sought to establish that the foundation of ancient Greece, and later the West, was significantly based on ancient Egypt's prototype. The only "twist," at least according to a few of Bernal's and Diop's detractors, is that both writers had no qualms about attributing ancient Egypt's rulers and populations to being Negroid. This ultimately gets to the heart of the debate, between the Afro-centrists and Euro-centrists. If Bernal and Diop are accurate in their independent assessments, then Western civilization owes its

75

colorful past to black Africans, and not the Greek and Roman civilizations, as proven racist scholars and intellectuals at the height of global imperialism in the late 17th, 18th, and 19th centuries, sought to impose their individual, collective, and national agendas on the conquered, less-fortunate majority of the world. According to Bernal and Diop (who would later be joined in their assessments by G. M. James George in his *Stolen Legacy* [2001]), the Greeks and Romans were mere conduits between an African past and a multiethnic present and future. George takes it even further and claims that the West's African past, some of which is contained in the teachings of Egypt's mysteries, took a circuitous path and found its way over many centuries to other societies, influencing their development *prior* to ever reaching Greece.

To refute Bernal's, Diop's, and George's assertions, many traditional classicists supported Mary Lefkowitz, at Wellesley College, and her colleague Guy Maclean Rogers in their attempt to provide a suitable rebuttal in *Black Athena Revisited* (1996). They were the editors of this volume, which contained 20 essays written by fellow colleagues, with the sole purpose of stripping Bernal's *Black Athena* of all its core arguments. A few of the essays succeeded, but as a whole, the book failed in its overall attempt and in some cases bolstered many of the important points made by Bernal. Lefkowitz, the primary editor of *Black Athena Revisited*, along with another book that she solely authored in 1996, *Not Out of Africa: How Afrocentrism Became an Excuse to Teach Myth as History*,

---

The Kingdom of Kush and Ancient Egypt

Lying just south of the first cataract of the Nile River was the Kingdom of Kush, a black African kingdom that exerted a large influence at a formative time in the history of the Egypt, before the emergence of what would become known as the New Kingdom. During 2006–2007, archaeologists staged a frenzied excavation of Kushite sites along the Nile in northern Sudan, sites that were soon to be submerged when the Merowe Dam was completed. The results of the expeditions showed that Kush had a much wider influence in the northern Africa region than was previously thought. The excavation of literally thousands of sites has revealed that Kush was much wealthier and more influential than previously thought; during approximately 2000 BCE to 1500 BCE, Kush developed a politically and culturally powerful kingdom and grew wealthy through gold refining and a powerful, centralized government despite the absence of a written language. Although it is difficult to argue that direct ties exist between the Kushite culture and later Egyptian culture, the indirect influence is easy to see, as the ancient Egyptians were well aware of Kush's power, since most of the written records on Kush come from Egyptian sources.

Source: Wilford, John Noble. "Scholars Race to Recover a Lost Kingdom on the Nile." *New York Times*, June 19, 2007.

The largest site of Kush civilization's burial pyramids is at Meroë, Sudan. (Uros Ravbar/ Dreamstime.com)

adopted a shrill approach in her attack against the Afro-centrists. This was evident not only in the deep urgency she felt that other scholars and lay readers needed to disregard in toto what was written by Bernal, Diop, George, and their ilk, but that they "distorted" the initial assumption that European culture was the ultimate be all and end all final destination for all other developing societies. Afro-centrists, according to Lefkowitz and her fellow collaborators, kept their sole focus on African civilizations, with no appreciation for what other civilizations offered and continue to offer the world. Lefkowitz not only felt that Afro-centrism was an attack on the central tenets of scholarship, she went as far in *Not Out of Africa* to demand that university administrations, throughout the nation and by extension the world, recognize the "truth" (a clear red flag and misnomer to most conscientious practicing historians, to say the least) in her research. According to Lefkowitz, administrations needed to set academic standards and rein in aberrant types as Bernal, Diop, and George, for they and others like them would ultimately contribute to the demise of democracy.

Besides her later emotional appeal to all academic administrations, Lefkowitz, in *Not Out of Africa*, solidly refuted Bernal's and Diop's contentions that Socrates and Cleopatra were black Africans and that the Egyptians and North Africans were not of Ethiopian lineage. She also stated that Egypt and their earlier conquerors, the Hyksos, never invaded, let alone conquered, Greece. Her argument begins to falter as she attempts to unconvincingly, without the requisite archaeological works, documents, and facts, have her readers believe that the Hyksos could

not have conquered Greece from Egypt, since they initially came from Greece. Obviously she forgot to keep her focus close to the till. It's a given among all classicists that the Hyksos were most likely of Syrian-Palestinian origin. Also, prior to the Thera eruption, that early Minoan civilization was actually Greek. Many scholars continue to debate that important point.

Lefkowitz also takes on George's teachings of Egypt's mysteries. She claimed that most of the purported Egyptian mysteries, with their philosophical and mystical bent, which Bernal and George claimed in each of their works as being ultimately adopted by the early ancient Greek sources, had no element of truth to them. Her flimsy response was that ancient Greeks often exaggerated in their writings; and for lack of confidence in their innate talents and history, they emulated and invented Egyptian sources as a means to *legitimize* their crafts to a widespread audience. According to Lefkowitz, by falling under the spell of Egypt's more stable and impressive history and culture, many of those Greek thinkers and writers would in time develop their own unique societal path, which would spew forth the artistic, philosophical, and political institutions that Europe, and later the United States, would identify the early Greeks with and adopt within their own political institutions and cultures. But, that was exactly the point that Bernal and George had made. Greek thoughts and ideas, the sort we associate with Socrates, Plato, Aristotle, and Pericles, to mention a few, got their origin from Egypt and North Africa. Most, if not all, of these well-regarded and recognized Greek individuals, at one time or another, spent some time in Egypt and other North African locales in their travels and endless quests for knowledge. With Alexander's conquest of Egypt and establishment of the city to be named after him (he would often send to his teacher, mentor, and original naturalist, Aristotle, numerous specimens to engage his entomologic interests), the city of Alexandria emerged as the most prosperous Greek community outside of Greece. The Greek community in Alexandria became for much of Greek history, up until the mid-20th century, the fountainhead and the most influential community in internal Greek politics and affairs. To most Greeks on the mainland and throughout the archipelago of islands and on Cyprus, the Greeks in Alexandria, with their wealth, culture, and fashion, remained the ideal to which they continued to strive.

In her raging rebuttal, Lefkowitz addressed the well-established Masonic traditions in Europe and the United States. Such traditions were always accepted by scholars and its devotees, throughout the ages, as organically linked to a system that could only have deep roots in the Egyptian mysteries. By a subtle and often not so subtle slight-of-hand, without any solid evidence to the contrary, she dismissed Egypt and ancient Egyptians' critical and central role in the continuing existence of such traditions in European and American Masonic orders. Her idea was that the sources and existing literature, which were Hermetic, and the initiation methods used by the Orders, which were shrouded in

secrecy, were derived completely from the Greeks. Quite an astounding claim, when you see that numerous well-respected Egyptologists, such as S. Morenz and T. G. H. James, despite their separate claims that there was nothing mystical or uniform about the religion of the ancient Egyptians, strongly held to the existence of an extensive body of knowledge that was passed down from one generation to the next. This knowledge, and its ultimate acquisition by Order members, Morenz and James believed, caused representatives within these Masonic Orders to adopt strict rules over the centuries for their initiation rites. Finally, Lefkowitz brushed aside the significance of Egyptian hieroglyphs as being primarily letters that had no other significance. A totally absurd notion, since from all archaeological and documented evidence from the Late Egyptian period, it's quite clear that the ancient Egyptians used hieroglyphs, not merely as writing tools, but, most inscriptions found throughout the temples were infused with an enigmatic nature. This was the common practice instituted in the Egyptian temples for millennia prior to the Late Egyptian period.

This, therefore, brings the reader to the reality of what resides at the center of the ongoing conflict between the Afro-centrists and Euro-centrists. Simply put, it is racism. Racism has reared its ugly head and will continue to rear its insidious face within academia's noble mission. This mission is to seek knowledge with no preconceived or stereotypical notions, with the duty to follow the evidence and facts to wherever they might lead. With that said, then, it could provide one the requisite liberty to investigate the intrusive and destructive patterns of racism, which as free thinkers we need to avoid, that's if, as scholars and serious thinkers, we wish to get on with our work. It should be quite clear to most readers that racism has even penetrated archaeological and documented historical evidence that *initially* was retrieved in a neutral manner at their sites. This has sadly, especially in archaeology, been its existential state as an exact school of inquiry since its inception.

As earlier mentioned, at the height of European imperialism and colonialism in the late 17th, 18th, and 19th centuries, the concept and later institution of race in the religious, political, social, and cultural spheres took hold. It was used to legitimize white European supremacy over all conquered people throughout the world. Its insidious nature would force it to justify its presence by attempting to skew the biological facts available to most, if not all, academic practitioners. By doing so, classified variations between the conquered people and other races were created; all, of course, for the "greater" good. Categories, subgroups, and subgroups to the subgroups were diligently created. They created and classified types of hair, skin, skull shapes, blood, eye, nose, and so forth. In time, variations were not only limited to those hair-splitting details, but they began to cast negative dispersions on the intellectual and moral basis and skills of the subordinate minorities within the European powers borders, but, more specifically, on the conquered natives within the borders of their extensive empires. This "scientific

racism" of the mid- to late-19th century categorized different races, such as the Caucasian race, the Australoid race, the African or Negroid race, and the Mongoloid race, into cozy and neat terms. The primary motive behind this method of classification was, once again, to prove beyond a shadow of a doubt the superiority of the Caucasian brain size and mind over those of all the other listed races. It would insidiously remain, despite unanimous rejection of its tenets by future generations of scholars in the humanities and social sciences in the 20th and 21st centuries (with positive scientific proof that it was impossible to isolate humans solely by biological traits and characteristics, which would predetermine their individual or collective responses), lurking at the subterranean mental and emotional levels with in the academic community and the extended informed and uninformed citizenry alike within the larger society. As most television, radio, Internet, and iPod users witnessed during the heated Democratic Party primary race between Barack Obama and Hillary Clinton in 2008, her campaign managers *knew* that if they were to blow on that inaudible "dog whistle" of white working-class fears of a black man, or other ethnic type, to win the coveted office of the U.S. presidency, then some members of that class, or even a few from the middle and upper classes, would respond to that sinister appeal by either voting against the black candidate or by not giving in to their party's requests or demands that they help the designated Democratic Party's candidate win the general election. Why would Lefkowitz title her book *Not Out of Africa*? Was that the inaudible "dog whistle" that she also *knew* could get a fearful and usually informed Caucasian audience to sit up straight and take notice? After all, many of her arguments, as earlier demonstrated, had no viable scholarly legs to stand on, except to appeal to the instinctual and emotional nature of a captive reading audience, who could see their own classical roots and history were being questioned. In time, if the archaeological evidence being gleaned supported a different version of history (instead of endowing it with the accurate assessment that the history being investigated was theirs and not some "foreign" body of work), then the one they had grown accustomed to and comfortable with would be discarded. That recipe for historical investigation, instead of becoming a groundswell for solid scientific archaeological inquiry, became a debased ground for the creation of mush.

In time, with the advent of the 1980s (a few short years *prior* to Bernal's published first *Black Athena* in 1987 and Lefkowitz's *Not Out of Africa* in 1996), a new form of cultural racism replaced the old. This new racism based its premise on essentialist and cultural grounds. Much more subtle than the earlier 17th, 18th, and 19th centuries variety, it manifested itself by placing much more emphasis on the "differences" in cultural heritage and traditions, with emphasis on the distinct nature of each of the groups being studied, measured, and discussed. This could be associated with the rise of nationalism and the strong presence of nationalist and ethnic group identity issues, especially after a volatile period of independence and anticolonialism in the preceding three decades

throughout the world. That was especially the case in a newly independent and struggling Africa. This could cast a clearer light on the modern inclination of many natives of Western societies to be less accepting of new immigrants within their borders, aside from some of the economic realities that also contributed to such an assessment—whether within America and some of the usual xenophobia expressed toward nonregistered illegal aliens, or French intolerance toward their newly arrived immigrants from the Arab world, North Africa, and West Africa. This, of course, would include some Dutch negative responses, as with the French, to their local minority populations, as evidenced with Arabs, Muslims, and South Asians, or in Great Britain and Germany, toward first-, second-, and third-generation Muslims of Turkish, Arab, Pakistani, Indian, and African heritage. Later, during George W. Bush's "War on Terror" era, it was easy for these European societies and the United States to use the broad brush stroke of cultural exclusion, which in many cases, unsuspectingly by those nations leaders, helped stoke the nationalist movements within these diverse societies: nations such as Iraq, later Afghanistan, and to a lesser degree, in Pakistan. Ethnic politics would also make a decisive resurgence in most of those regions of the world. That can currently be witnessed in Nigeria, Rwanda, Sudan, Zimbabwe, and Kenya. It's within these diverse religious, political, economic, social, and cultural settings that a sense of "superiority" arose, which was based on racial factors and not perpetrated *solely* by the West against those countries and their populations. Rather, it was a racism from *within* these disadvantaged regions of the world, which was not only directed to the West and outside world, but, more tragically, to the other races and ethnic groups within each nation identified with their respective regions. To say the least, it was and continues to be quite a sorry state. Exacting revenge and innate pride in the historical and cultural values of one's defined tribal group politicized archaeological sites; instead of weaving them into a professional and intellectual undertaking, it wove them into a symbolic and charged cultural setting. That was what we witnessed to a certain extent with the volley of accusations and counteraccusations emanating from Martin Bernal, Cheik Diop, G. M. James George, and Mary Lefkowitz and Guy Maclean Rogers. With the culturally charged academic settings, neither side felt the need to reconcile its strongly held beliefs. Sadly, it continues to be so.

In conclusion, perhaps the last word on the matter as to whether Classic culture has Afroasiatic roots (not that that's ever intellectually viable in most research and academic related matters) could be found in those whose ancestors are at the very center of the storm—the Egyptians themselves. With the backdrop of the recent discovery of pharaonic temples from 1512 BCE, from the reign of Thutmosis II, in the Sinai Peninsula in early April 2009 on many Egyptians' minds, many were asked their opinions of the theory, including colleagues and scholars, white-collar and working-class men and women, farmers,

along with the usual sample of random responses by the average Egyptian man and woman in public gatherings. The general consensus, after a very long hearty laugh, was "Who cares?" One elderly man wisely said: "If my great-great grandfather and mother were black, and they achieved these wonderful things, I'd be just as proud of them, as I would be if they were white, or any other color. It's a human thing." A chorus of fellow Egyptians stood by and nodded their complete approval. Instead of Afro-centrism and Euro-centrism, the Egyptians captured the essence of world-centrism—"It's a human thing." It was a very practical response by a homogeneous people to some members of the heterogeneous societies who continue to seemingly be beating a dead horse. In that typical Egyptian response, possibly "*M'isr, 'Umm al-D'unya*" (colloquial variation often used by Egyptians when they refer to their society's role in history and the world, always with a chuckle and smile, meaning "Egypt, the Mother of the World"), there were a few more lessons for the Bernals, Diops, and Lefkowitzes of the world. That message, in a hopefully more postracial world, is "Lighten up. Use your common sense. Do good research."

References and Further Reading

Baker, J. *Race*. Oxford: Oxford University Press, 1974.

Bernal, Martin. *Black Athena: Afroasiatic Roots of Classical Civilization: The Archaeological and Documentary Evidence*, Vol. 2. Piscataway, NJ: Rutgers University Press, 1991.

Bernal, Martin. *Black Athena: Afroasiatic Roots of Classical Civilization: The Fabrication of Ancient Greece, 1785–1985*. Piscataway, NJ: Rutgers University Press, 1987.

Bernal, Martin. *Black Athena: Afroasiatic Roots of Classical Civilization: The Linguistic Evidence*, Vol. 3. Piscataway, NJ: Rutgers University Press, 2006.

Corrado, G. "Psychic and Cultural Traits and the Classification of Human Races." *Mankind Quarterly* 1 (1962): 236–41.

Dennett, Daniel C. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Schuster, 1995.

Diop, Cheik Anta. *The African Origin of Civilization: Myth or Reality*. Chicago: Lawrence Hill, 1974.

Diop, Cheik Anta. *Civilization or Barbarism: An Authentic Anthropology*. Chicago: Lawrence Hill, 1991.

Ekman, Paul. *Emotions Revealed*. New York: Henry Holt, 2003.

George, G. M. James. *Stolen Legacy*. New York: African-American Images, 2001.

Lederach, John Paul. *The Moral Imagination: The Art and Soul of Building Peace*. New York: Oxford University Press, 2005.

Lefkowitz, Mary R. *Not Out of Africa: How Afrocentrism Became an Excuse to Teach Myth as History*. New York: Basic Books, 1996.

Lefkowitz, Mary R., and Guy Maclean Rogers, ed. *Black Athena Revisited*. Chapel Hill: University of North Carolina Press, 1996.

Nasr, Seyyed Hossein. *Knowledge and the Sacred*. Albany: State University of New York Press, 1989.

Wilson, David Sloan. *Darwin's Cathedral: Evolution, Religion, and the Nature of Society*. Chicago: University of Chicago Press, 2002.

CON

The main thrust of Martin Bernal's culture theory, that much of early Classic Western culture has its origins in Africa and Asia, is ideologically motivated, with factual evidence in short currency. From the contents of his works, it becomes clear that the attribute "black" in the main title (i.e., *Black Athena*) is synonymous with "ancient Egyptian" rather than referring to black in the meaning of African black.

Undoubtedly, the roots of various Egyptian institutions and cultural traditions (e.g., the worldview of the ruler of divine descent, the pharaoh, as a reflection of African sacred kingship) are to be sought in the prehistory of the black African populations. On the other hand, perhaps more cultural patterns are of local Egyptian coinage and developed without influence from the heritage of black African cultures (e.g., writing technology, monumental architecture, the creation of new gods and goddesses). It is noteworthy that the black Nubians who lived in the south of Egypt learned the crafts of writing and of pyramid building from the Egyptians, and not vice versa.

A significant marker of distinction, and separation, of the ancient Egyptians from the black populations of Africa is language. Ancient Egyptian is affiliated with the Afroasian (Afroasiatic) phylum and unrelated to the languages of the black populations that belong to other language families (i.e., Niger-Congo, Nilo-Saharan). Since language is a prominent vehicle from which to construct culture and knowledge, ancient Egyptian civilization created its own specific blend of cultural heritage.

Bernal intentionally chose a provocative title for his work to stir up the Europeans who, in their Euro-centric bias, had created a fanciful image of Greek antiquity and the origins of its civilization. In the zeitgeist of nationalism of the 19th century, the concept of "national culture" was idealized as an independent entity, and high esteem was given to those cultures with a long history, along the lines of "the older and purer, the more precious." National cultures in Europe

were categorized accordingly and, when it came to the evaluation of Greek civilization, a fabric of idealized Hellenism emerged, glorifying Greek ingenuity and hailing the purity and originality of Greek cultural traditions.

This construct of pure and original Greek civilization was more a reflection of intellectual preoccupations of the national cultures of Europe rather than a factual reconstruction of life in antiquity. In fact, this fabric was so attractive that many writers, poets, and philosophers preferred to live with their fanciful dreams rather than get acquainted with the reality of Greek traditions (Lambropoulos 1993: 57).

Bernal's attack is directed against the self-contained way in which the image of ancient Greek civilization was distorted by European cultural chauvinism. Calling those who advocated Euro-centrism "racists"—as Bernal (1987: 201ff.) does—means barking up the wrong tree. The Euro-centrics who hailed the purity of Greek civilization applied the same categories to the cultures of other "old nations" like the Germans, French, English, or Russians. Since Bernal's intention is to provoke, he applies the recipe to fight the devil with Satan, and he advocates the opposite extreme of what he stands up against. According to this logic anyone who is not an Afro-centric is a racist. With this distortion as a guideline Bernal crosses the line that divides serious scholarship from ideology. As Guy MacLean Rogers argued:

> Bernal's main strategy . . . has invited, and perhaps even insures, however, the opposite of his intended effect of lessening European cultural arrogance. He pays attention and gives stature to Egypt and Phoenicia only in relation to a later and different Western culture. Thus *Black Athena* succumbs to exactly the Eurocentrism it was written to combat. (1996: 452)

The enlightened scholars of classical antiquity, starting with Gottfried Wilhelm Leibniz in the early 18th century, have never negated Near Eastern and Egyptian influences on Greek culture and language, but critical scholarship has always known to make a distinction between cultural exchange among equal partners and the foundational impact of one culture on the formation of another. That means that critical scholarship has kept a balance and avoided extreme positions because neither the purely Euro-centric nor Bernal's Afro-centric doctrine can serve as a viable orientation for explaining the fabric of Greek civilization.

Arguably, most students of Greek antiquity of the 19th century were caught in the Euro-centric zeitgeist. Speaking of zeitgeist, this holds true for the position of Bernal himself. The elaboration of his Afro-centric theory of classical culture falls in a period (i.e., the late 1970s and early 1980s) when fundamental changes in the academic world of archaeology and culture studies were under way. One major innovation in archaeology was the application of dendro-chronological dating methods (i.e., tree-ring dating) to calibrate older radiocarbon (i.e., C 14) dates for reaching reliable connections with absolute time. The new

cultural chronology for European prehistory was something of a revolution. What had been thought of as spanning a time of some hundreds of years turned out to have lasted several thousand years. The new chronology opened the view into the depth of prehistory that had previously been blurred.

Archaeologists learned that agriculture spread to southeastern Europe from western Asia already around 7000 BCE, some two thousand years earlier than the beginnings of food production in the Nile Valley. Marija Gimbutas (1982, 1991) deserves the credit for mapping out the civilization of the early (pre-Greek) agriculturalists of Europe, for whose regional cultures she used the blanket term "Old Europe." Old Europe (or the Danube civilization, as this complex is referred to in recent studies on archaeology) developed the foundational institutions of high culture at a time when Egyptian civilization as we know it did not yet exist.

Among the markers of the Danube civilization that flourished from ca. 5500 BCE to ca. 3500 BCE were an extended network of trade relations, a sophisticated belief system (with refined religious concepts and practices centering on a supreme female divinity), and advanced technologies such as pottery making, weaving, and metal working (copper since the sixth millennium BCE and gold since the fifth millennium BCE). The web of communication systems was highly varified, with vivid cultural symbolism and notational systems. One system served for rendering numbers; the other was an archaic form of writing, referred to as "Old European script" (Haarmann 1995) and also as "Danube script" in recent scholarly literature (Marler 2008).

This pre-Greek (pre-Indo-European) civilization declined under the impact of Indo-European culture whose bearers came from the steppes of southern Russia and intruded into southeastern Europe in the fourth millennium BCE (Anthony 2007). Old European traditions were gradually overformed on the mainland, but they reached the islands in the Aegean Sea in a cultural drift, giving rise to the Minoan civilization of ancient Crete (of the third and second millennia BCE), which, in turn, decisively shaped the profile of Mycenaean-Greek culture of the second millennium BCE.

From the standpoint of the newly established cultural chronology of southeastern Europe, the foundational institutions of which classical Greek civilization is comprised find their origins in cultural layers that predate the rise of civilizations in ancient Egypt and in Mesopotamia. Giving emphasis to the indigenous (European) roots of pre-Greek civilization does not stand in opposition to considerations about Afroasiatic influences in Greek culture. What is crucial for any assessment of such influences is the proper correlation of the phases of Egyptian and Mesopotamian influence with the cultural chronology of Europe. Bernal fails to establish such a correlation, and he does not use the pertinent scholarly literature that informs about Old Europe and the achievements of the pre-Greek population.

Already early in the 19th century there was an awareness among open-minded academic people in Europe that there had been a constant seepage of

Egyptian and Near Eastern influence into Greek culture. The evidence for such influence dates to the second millennium BCE. The knowledge of such influences appears in a new light when assigned its proper place in the cultural chronology.

Since the times of the New Kingdom in Egypt's dynastic history (ca. 1550–1295 BCE), Egyptian cultural influence radiates far from the heartland into other Mediterranean civilizations. Wherever and in whatever form the Egyptian influence is manifested in the Greek world, it is received by a culture that had grown from ancient roots already at a time before the diffusion of Egyptian cultural impulses. This is true for the Minoan culture in Crete and for the Mycenaean-Greek culture on the mainland. Greek civilization of the archaic and classical period (eighth to fourth centuries BCE) is an offspring of traditions that reach back in time—via the ancient Aegean cultures—to the Danube civilization, which makes Greek civilization the grandchild of Old Europe. Its fabric is that of a true conglomerate or mosaic, with a fusion of indigenous and "imported" constituents.

The scenario of how this mosaic culture came into being will be highlighted in a variational approach by illustrating the infrastructure of classical Greek culture and its various constituents rather than by reflecting exclusively on the single concept "roots."

Language as an Identifier of Pre-Greek Civilization

The pre-Indo-European population that inhabited the areas around the Aegean Sea and its islands in pre-Greek times left a genetic "footprint." Geneticists call the genomic profile of that ancient population the "Mediterranean outlier" because it differs markedly from genetic configurations in surrounding areas (see Budja [2005] and Haarmann [2007] for a reconciliation of data from human genetics and archaeology). The ancient population of Crete, the Minoans, also formed part of this area. In some speculative approaches, the Minoans are seen as descendants from traders who allegedly came to Crete from North Africa in prehistoric times. However, human genetics cannot verify such claims. There is no genetic linkage between the ancient Cretans and North Africa.

The Old Europeans, bearers of the Danube civilization, left other traces, and these are the linguistic remnants of their language that survive in Greek as a substratum (i.e., an underlying layer) in word formation and in the vocabulary, and from the structure of the linguistic material it is clear that the pre-Greek language was unrelated to any African phylum.

The pre-Greek elements can be identified in the names of places, rivers, and formations of the natural environment. Characteristic of the roots of names of pre-Indo-European origin are certain formatives (i.e., -ss-, -nd-, -nth-) of which -ss- is the most frequent in this group of names: for example, Assa, Larissa, and Passa in the Balkan region, Bubassos, Sardessos, and Termessos in western Asia

(Asia Minor). In the Aegean archipelago and in the Balkans we find altogether 181 names in contrast to 175 names in Asia Minor.

It is noteworthy that the pre-Greek names include many designations for towns and cities.

> From this fact can be concluded that the speakers of Greek found at their arrival the Aegean urbanization already in full development. This observation fits well with the inference from archaeological evidence according to which the first peak of urban life in the Aegean World was reached already in the Early Bronze Age, whereas Greek-speaking tribes invaded the area only at the beginning of the Middle Bronze Age. (Katičić 1976: 55)

The pre-Indo-European language left traces, in manifold transformations, in the lexical layers of ancient Greek. Hundreds of terms in various domains of the vocabulary were adopted as elements of the pre-Greek substratum. Some examples of pre-Greek loan words are listed here, focusing in particular on those elements that document the high level of cultural development among the pre-Greek population:

Names for plants: *ampelos* "vine," *aroma* "aromatic plant," *daphne* "laurel," *elaia* "olive tree," *ion* "violet," *minthe/minthos* "mint," *narkissos* "narcissus"

Textiles, clothing: *beudos* "precious female dress," *eanos* "long female dress," *tebenna* "cloak-like aristocratic dress"

Structures, living space: *gefura* "bridge," *ikrion* "board, plank," *klibanos* "bread-oven," *asaminthos* "bathing-tub"

Handicraft: *keramos* "clay (for producing pottery), pot," *skiros* "gypsum," *elakate* "spindle"

Food, dishes: *eia* "flour, clump of grain," *thargelos* "bread which is baked from the first grain of the harvest," *itrion* "bread with sesame"

Religious symbolism and practices, mythology: *bretas* "effigy made of wood," *thiasos* "procession in a religious ceremony," *thriambos* "ritual accompanied by chanting and dancing," *megaron* "sacrificial pit; holy precinct"

Social relations: *anassa* "mistress (female ruler)," *karbanos* "barbarian, stranger," *thes* "hired worker"

Adjectives: *afanros* "fragile, weak," *katharos* "pure," *khalepos* "difficult," *kednos* "venerated"

Verbs: *blepo* "gaze, see," *brizo* "sleep," *dunamai* "be mighty, potent," *iapto* "throw, sling"

The occurrence of verbs and adjectives in the repertory of borrowings points at the fact that the contacts between Greeks and the pre-Greek autochthonous

populations were intense. The quality of pre-Indo-European borrowings differs markedly from hunter-gatherer vocabulary. In fact, the substratum in ancient Greek stems from the language of a sedentary population with agrarian lifestyles and advanced technological skills. Among the borrowed names for plants are those for cultivated plants (e.g., *ampelos* or *elaia*). Some borrowed terms for textiles that became popular in Greek society (e.g., *tebenna*) point at special fabrics. In the domain of handicraft and technology, the Greeks learned many things from their skilled predecessors, namely how to build bridges (e.g., *gefura*), to use bread-ovens (e.g., *klibanos*), and to practice weaving (e.g., *elakate*).

Traces of the ancient terminology of pyrotechnology and metallurgy have been preserved in the vocabulary of ancient Greek. Pre-Greek origin is assumed for the following key terms: *kaminos* "furnace," *kibdos* "metal slag," *kassiteros* "tin," *metallon* "metal," *khalkos* "copper." The latter term is associated with *khalke/kalkhe* "purple snail," which is also a pre-Greek borrowing. The name for copper in Greek, thus, is derived from the reddish color of the metal.

The religious traditions of the pre-Indo-European communities in southeastern Europe must have impressed the Indo-European intruders because a number of key concepts of their religious life entered the ancient Greek lexicon (e.g., *bretas, megaron, thiasos*). If the indigenous population in the Balkanic region had been simply replaced by the Indo-European migrants, then hardly any such expressions would have been adopted. Moreover, terms such as *thiasos* and especially *megaron* not only retained their original (pre-Greek) importance but their significance for Greek religious life even expanded during the classical period.

Bernal dedicates the third volume of *Black Athena* (2006) to the discussion of the linguistic impact of Egyptian on the Greek language, which, according to the Afro-centric view, is allegedly massive. The extensive survey of reconstructed Egyptian roots for Greek expressions, however, lacks an "anchor." Bernal cannot explain how hundreds of borrowings would have entered the Greek vocabulary without any well-established interaction between the people who spoke the languages in contact. While the presence of the elements of the pre-Greek substratum in the Greek lexicon can be explained through social interaction and cultural exchange between the Greeks and the local agrarian population living in the region that was to become their homeland, the alleged flow of Egyptian terms is not anchored for the simple reason that there was no intensive interaction between Egyptians and the populations of Europe in the third millennium BCE, which, according to Bernal's claims, is the decisive time frame for massive borrowing.

Serious scholarship does not negate the presence of Semitic and Egyptian loanwords in Greek. However, less than 30 words are widely acknowledged to have been adopted from Semitic sources, and the transfer of Egyptian borrowings into Greek is of a relatively late date. As Jasanoff and Nussbaum put it: "The Egyptian words in Greek are on the whole fewer and later than the Semitic words. A high percentage are confined to the Greek spoken in Egypt in

Hellenistic and Roman times; others are merely quoted as foreign words by late Greek authors" (1996: 188). Most of these late borrowings are elements that enriched the cultural vocabulary of the Greek language (e.g., Greek *papuros* "papyrus," *ibis* "ibis bird," *kommi* "gum"), but they do not form part of the basic vocabulary that would be a precondition for demonstrating the validity of the Afro-centric theory.

The Prominent Women of Greek Mythology and Their Pre-Greek Ancestresses

Since Bernal associates his Afro-centric speculations with the figure of Athena, some reflections on this goddess are called for. According to Bernal's claim, the Greek goddess Athena is a figure that was coined on an Egyptian model, and he gives several potential "candidates" of Egyptian goddesses that allegedly inspired the creation of the Greek Athena:

- The goddess Nephthys (Bernal 1991: 99*f*.). This juxtaposition is difficult to perceive since a comparison of the two goddesses reveals quite differing functions for each. The Egyptian goddess was married and had a child with Osiris, while Athena was venerated as the eternal virgin who sought the company of men but never got involved with them. Nephthys is the goddess of the realm of the dead and the patron of birth giving, both functions that Athena never had.
- The goddess Neit (Bernal 2006: 563*f*.). Neit was the local patron of Sais in the Nile Delta. She was venerated as a goddess of the river and the sea. As a primordial goddess, Neit is the mother of the sun god Ra. As a goddess of the underworld Neit shares important functions with Nephthys. In the late phase of pharaonic Egypt, Neit becomes the patron of kingship who bestows the regalia upon the pharaoh. All these functions lack resemblance with Athena and her role in Greek mythology.

Athena is neither of Egyptian nor of black African origin and, being the daughter of the prehistoric goddess of the Old Europeans, she is as old as any of the Egyptian goddesses. In order to understand Athena's role in the web of mythological relations, it is necessary to illustrate the position of Greek goddesses in the Greek pantheon ("circle of divinities"). A closer inspection reveals that female divinities are the most ancient figures of Greek mythology and date to pre-Greek times. A goddess is even at the very heart of the Greek myth of origin. The functional prominence that female divinities enjoy in Greek mythology is unknown to the Egyptian tradition, which, therefore, cannot be seen as the source of the former.

The mainstream Greek cosmogonic myths (i.e., myths of origin) are collected in the epic work "Theogony," composed by the Greek poet Hesiod

around 700 BCE, although there are other cosmogonies categorized as "deviant." Among the "deviant" cosmogonies is the so-called Pelasgian myth of pre-Greek origin that is of high age. According to this version, the first divinity to emerge from the primordial chaos is Eurynome ("wide wandering"), the goddess of all things, whose first task is to divide heaven from the eternal waters. She cannot find a place to rest, so she wanders south on the waves.

Eurynome is followed by the north wind, Boreas, who takes the form of a snake. His longing for a female companion culminates in a sexual union with the goddess who becomes pregnant. Eurynome experiences a metamorphosis, turns into a waterfowl, and places an egg on a patch of land. From this egg emerge all the things of this world, living and nonliving. In this version, the central position of the goddess as creator of the world is indicative of the pre-Greek origin of the cosmogonic myth. Moreover, Eurynome originally presided over Olympus before she was dethroned by Zeus and the other Greek Olympians.

With respect to the cultures of the Aegean islands, the pre-Indo-European heritage of the goddess cult is best known from ancient Crete with its Minoan civilization that flourished in the second millennium BCE. The opinions of scholars are divided over the fabric of Minoan religion. Some assume the presence of one mighty female divinity, while others reconstruct a pantheon of gods and goddesses. Even if Minoan religion knew various divinities, the prominence of female deities among them remains striking. "That a powerful goddess of nature was the chief deity of the Minoans . . . has never been seriously questioned" (Marinatos 1993: 147).

The heritage of the "strong women" in pre-Greek mythology is reflected in the goddess cults of ancient Greece. All major achievements of Greek civilization are associated with goddesses (Haarmann and Marler 2008: 48*f.*):

- Demeter—the patron of agriculture (the Grain Mother);
- Hestia—the guardian of the hearth and the household;
- Artemis—the patron of nature and of wildlife;
- Athena—the supergoddess, the patron of technologies such as pottery (the potter's icon was Athena's owl), shipbuilding and weaving, the patron of justice, of the arts and science. As a protectress, Athena watched over the safety of Athens and the well-being of its citizens. The main temple on the acropolis in Athens, the Parthenon, was dedicated to her.

Evidence for the fact that Athena is of old age and that this figure originated in pre-Greek times can be found in the name itself. The formative element *-n-* (as in the last syllable *-na*) is not typical of Greek word formation, and this element—as part of the name—was adopted by the Greeks from the language of the Old Europeans, together with the figure of the goddess and her cult. The information

that is available about the functions of Athena and the history of her cult practices makes any claim about Afroasiatic origins void.

Basic conceptualizations about fertility in a spirited world, related to the soil that produces crops to animals that were raised as cattle and to women as givers of life, persisted in many agrarian communities into the times of classical antiquity. In some religious traditions, the mindset of early sedentarists still echoes. This is true for the cult of the Greek goddess Demeter, the Grain Mother, and her festivities.

A clear reference to Demeter in her association with the cultivation of grain is given by the poet Homer in his epic *Iliad* (XIII 322, XXI 76, V 500), which was compiled in the eighth century BCE. In the cult practices of Demeter, the pig (particularly the suckling pig) played a prominent role. When the crops were sown in autumn, the festival of Thesmophoria was held in honor of Demeter. The Thesmophoria "give an impression of extraordinary antiquity" (Burkert 1985: 13). During this festival, which was performed and attended by women only, young pigs, among other votive gifts, were offered to the goddess.

Another festival, called Skirophoria, was also associated with Demeter. Virgins received figurines of suckling pigs (Greek *skira*), and these were left after the ceremonies in the shrine dedicated to Demeter. In the rites of the Eleusinian mysteries, the pig was the basic symbol of purification.

The traditions of these strong women of Greek mythology lived on and later fused with the cult of the Virgin Mary, who absorbed the authority of the ancient goddesses and attracted their worshipers. Among those Greek divinities whose features were absorbed by Mary, Athena may be the most revered. "By the time the Parthenon of Athens was dedicated to Mary as Mother of God in the sixth century, she had taken on many of the images and honours of the ancient goddesses as well as moving into many of their temples" (Shearer 1996: 118). Of all the female deities that were once venerated in antiquity, the prominent women of Greek mythology live on in our cultural memory—often in their Roman equivalents (Demeter as Ceres, Athena as Minerva or Iustitia, Aphrodite as Venus, etc.)—as ingredients of Western civilization, with many allusions to them in the visual arts and in literature.

The Question of Continuity: From Minoan and Mycenaean-Greek Culture into Later Periods

The pre-Indo-European civilization did not disintegrate under the Indo-European overlay, but patterns of a selective continuity are successively derived from the ancient foundations. In fact, "the Old European sacred images and symbols were never totally uprooted; these most persistent features in human history were too deeply implanted in the psyche" (Gimbutas 1989: 318). The vivid continuity of Old European features in later periods has given rise to the

Mary Lefkowitz

Mary Lefkowitz emerged as one of the most formidable critics of *Black Athena*. Born on April 30, 1935, in New York City, she studied at Wellesley College and then Radcliffe College (now a part of Harvard University), where she completed her AM in 1959, and her doctorate in 1961, her thesis being on "A Study of First Personal Statements in Pindar." From 1969 until 1975 she was assistant professor of Greek and Latin, and for the next four years was professor of Greek and Latin before becoming the Andrew W. Mellon Professor in the Humanities at Wellesley College, until her retirement in 2005.

Much of Lefkowitz's initial work was on the role of women and heroines in Greek society, and she wrote *Heroines and Hysterics* (1981), followed by *The Lives of the Greek Poets* (1981), and then edited *Women's Life in Greece and Rome* (1982)—the latter book going into three editions. Many other books followed, with Lefkowitz arguing that storytelling in major literature throughout the ancient world handled the gods in a similar manner.

idea that the Minoan civilization of ancient Crete (third and second millennia BCE) and the Mycenaean-Greek civilization (second millennium BCE) were daughters of the Old European mother culture.

The Cultural Drift from the Mainland into the Aegean

The selective transformation of Old European traditions into Aegean patterns was a process of a repetitive continuity of pertinent features of the pre-Indo-European culture, rather than a fragmentization of the original entity. The continuity was repetitive in the sense that, after a time of disrupture and after a period of cultural instability, the pre-Indo-European canon repeated itself in the Aegean civilizations, and several of its major characteristics continued to be significant as constitutive elements of Cycladic, ancient Cretan (i.e., Minoan), and (Mycenaean-) Greek culture.

Among them are the prominent role of female deities, their practices and associated items (figurines); foundational motifs of cultural symbolism such as the snake, spiral, meander, swastika, the horns of consecration (*bucrania*); systems of visual communication (i.e., the notation of numbers and prealphabetic writing). For these constitutive elements of the ancient Aegean society, a pre-Indo-European heritage can be postulated and, in the horizon of time, they are perpetuated into the civilization of classical Greece (Haarmann 1995: 57ff.).

A more "visible" Old European tradition inspired the cultures of the ancient Aegean where writing emerged as resulting from idea diffusion in the Balkanic-Aegean cultural drift. The sign inventory of Minoan Linear A owes about half of

its signs to symbols of the older Danube script. The Minoans in turn inspired the Mycenaean Greeks to write their language in a variety of their script, Linear B.

Most probably, Minoan literacy also influenced the Greek genres of literature. Scholars have always wondered why the Greeks would choose such a complex form as the hexameter (i.e., the six-footed verse) for their early epic poems, Homer's *Iliad* and *Odyssey* in particular. The epic poems reflect the language of oral tradition, and the Greek hexameter "transcends in its length, variability and its capacity for articulation all other known verse forms of oral poetry" (Dihle 1994: 9). It has been argued that the reason for this may lie in the pre-Greek origin of this text structure. Some scholars believe that the Mycenaean Greeks adopted the hexameter from the Minoans and used it for their own language. Blok argued that "The Minoan origin of the hexameter could explain why epic diction is such an artificial combination of dialects and linguistic stages: since Minoan is not Greek, the Mykenaian (i.e. Greek) development of the epic . . . implies that the language had to be adapted to an external metrical structure" (1995: 188).

Greek Culture throughout the "Dark Ages"

The political supremacy of the Mycenaean city–states declined in the 12th century BCE. On the Greek mainland, a period of cultural disorientation seems to have begun, and literacy in Linear B was abandoned. During those times of social unrest and political instability, many Greeks migrated from the mainland to Cyprus, taking with them urban lifeways and their cultural traditions. The lack of literary sources and the scarcity of archaeological artifacts between the eleventh and ninth centuries BCE produced the puzzle of what modern scholars use to call the "dark ages." Even on the mainland, however, the decline of Mycenaean rule was not followed by a total disruption in cultural tradition. In the visual arts of archaic Greece (eighth century BCE), in their form, style, and motifs, the legacy of the Mycenaean era can be clearly perceived.

There were no "dark ages" in Cyprus, and cultural development there is characterized by a continuity of cultural forms, including writing. Contacts between Cyprus and the Greek mainland continued even after the decline of Mycenaean power. The contacts of the eleventh century BCE were of mutual advantage, with the transfer of Mycenaean traditions to the island (e.g., decoration of Cyprian pottery with motifs of the Late Helladic IIIC period such as pictures of warriors and Mycenaean-style weaponry) and with Cyprian stimulations of a mainland iron industry and the import of a Cyprian innovation, "something of a revolution in pottery, the formation of a very distinct style known as protogeometric" (Osborne 1996: 28). In the cultural environment of Cyprus, Mycenaean patterns did not cease to exist, persisting instead until well into the first millennium BCE. The remainders of Mycenaean cultural persistence in Cyprus can be traced for several centuries, and "it becomes evident that during the first millennium B.C. the

Cypriotes cherished their Mycenaean ancestry and tradition and preserved or revived elements of Mycenaean culture even at times of foreign-oriental domination in the island" (Karageorghis 1962: 77).

In the bilingual milieu of ancient Cyprian and Greek cultural interference, literacy in the Cypriot-Syllabic script evolved from the 11th century BCE onward, until it was challenged by the Greek alphabet in the sixth century BCE, eventually being replaced by it in the third century BCE. Cyprus was not only a striking exception to the stereotyping notion of the "dark ages" in the Mycenaean-Greek hemisphere, it also played an active part in the perpetuation of the old Aegean heritage and its proliferation in the Mediterranean world into the archaic period of the eighth to sixth centuries BCE.

The Tradition of the Bull Cult in the Mediterranean and the Originality of the Minoan Bull Games

The strings of cultural continuity that Gimbutas and other scholars have identified for European cultural history from the Neolithic to the Bronze Age and beyond seem to have been ignored by Bernal. His discussion of the role of the bull as a mythical animal in Mediterranean cultures reflects this lack of information. Bernal claims that, during the period between the documentation of bull horns in sanctuaries of Çatalhöyük in Anatolia of the seventh millennium BCE and the appearance of the bull motif in Minoan art (second millennium BCE), there is no evidence for continuity in the region (Bernal 1991: 166). In fact, there is ample evidence in the imagery of Old Europe (e.g., bull figurines, the picture of a bull's head carved on a bone plate, the motif of the horns of consecration). Bernal's evaluation of this imagined lacuna of visual evidence is that, during that time, Egyptian influence comes to bear and, ultimately, inspires the Cretan tradition of the bull games.

These bull games are documented in the famous bull-leaping fresco in a courtyard of the Minoan palace of Knossos. For over a hundred years, the scene depicted in this fresco has puzzled experts and laypersons alike. Some see the leaping as the act of an individual actor, separate from the other actors. In fact, few have thought of a choreography involving several actors, three to be precise as they are depicted in the fresco. If the actor who clings to the horns had the intention to perform a *salto mortale* and land on the bull's back, an accurate landing would hardly be possible, given the unpredictable and accidental motion of the bull's head.

It is important to keep in mind that the scene does not depict the action in a realistic manner. Instead, "the artistic depictions concentrate on the essence of the feat, which consists of outwitting the bull, rendering him almost ridiculous by turning his potency into a vehicle for display of human skill" (Marinatos 1993: 218*f*.). The participation of three human figures in the scene speaks for a
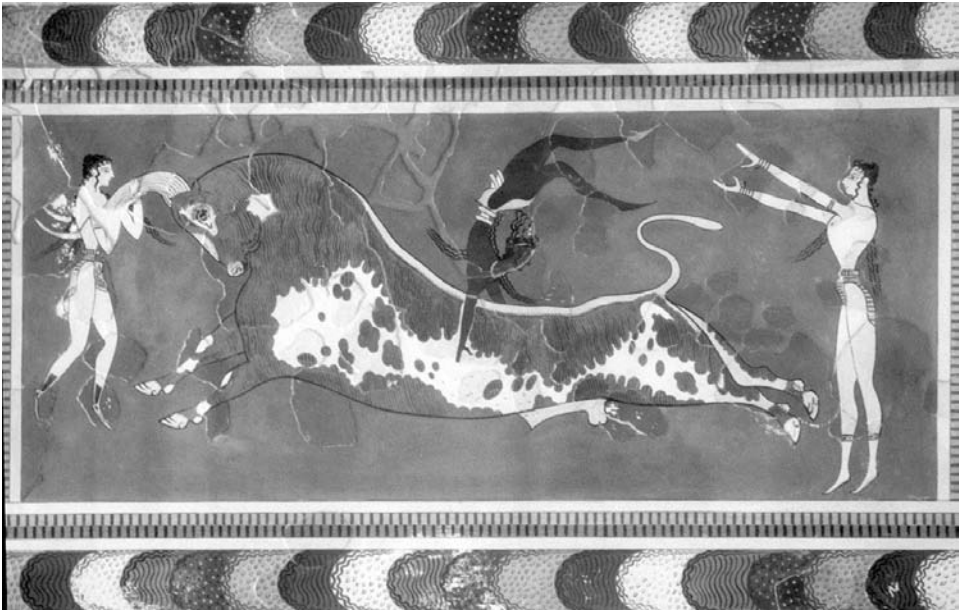
Fresco of a bull-leaping competition, from the ancient Minoan city of Knossos on the island of Crete. At the turn of the 20th century, British archaeologist Arthur Evans excavated Knossos, uncovering spectacular frescoes in the Bronze Age palace. (Corel)

reconstruction of bull leaping as a coordinated choreography, with several actors performing. The depiction of three acrobats (i.e., A, B, C) corresponds to a staged choreography. Acrobat A is the first to approach the bull, trying to hold its head down by clinging to its horns. (Such a motion is still typical of the Portuguese tradition of an unbloody bull fight.) Pressing down the bull's head is an important precautionary action, the purpose of which is to keep the dangerous sharp ends of the horns out of reach of the level where they could harm acrobat B who is the next to act. He waits for the right moment to leap over the bull's head while this is held down by acrobat A. Acrobat B performs an extended leap, lands on the back of the bull, and uses the thrust of the motion to jump over the animal to land on his feet. Acrobat C will assist him to land safely. (Leaping over a bull [or a cow] is a ceremonial action still performed during festivities in southern France.)

Once the leap has been successfully performed, acrobat A loosens his grip on the bull's horns so that the animal can run freely. The acrobats then gather in a slightly changed order. Acrobat B remains behind the bull while acrobat C takes the role of the one to cling to the bull's horns. Acrobat A, for her part, waits for a favorable moment to leap over the bull. Depending on the physical resources and skills of the acrobats, or on the requirements of the ritual, a choreography of this type could be performed in repetitive sequences for a shorter or longer duration of time.

In whatever way one may interpret the bull-leaping performance, this action that formed part of religious rituals was a Minoan invention, and there is nothing comparable in the mythological tradition of the bull and its sacrifice in ancient Egypt. The find of the picture of a leaping bull at Avaris in the eastern delta region, the capital of the Hyksos who ruled over Egypt during the second intermediate period (ca. 1650–1550 BCE), which shows resemblance with the Cretan fresco, cannot be overrated in its significance as a possible source for the Minoan tradition, as Bernal has it. Rather, this parallel documents to the vivid trade relations and cultural exchange between two civilizations, each with its own indigenous roots, capable of absorbing additional and manifold impulses from its neighbors.

Trade Relations versus Colonization in the Mediterranean Region

Trade goods of Egyptian origin are literally strewn in the areas around the Mediterranean Sea. In the archaeological record, one finds Egyptian miniature sculptures, among them the popular motif of the scarabee, the dung-rolling bug, which was considered a sacred symbol of eternity by the Egyptians. His obsession with Egyptian grandeur makes Bernal forcefully argue for Egyptian political power in the regions with Egyptian trade goods, or even for colonization. It is hazardous to exclusively interpret trade relations as an indicator of colonial power. Some examples may illustrate such a misconception. The archaeological record of the settlements near the southern coast of the Baltic Sea—in what is nowadays Polish and Lithuanian territory and dating to the early centuries of our era—shows an abundance of Roman coins. It would be absurd to interpret this concentration of wealth as an indicator of Roman colonial presence in the area. The coins stem from the vivid trade relations that the Romans entertained along the amber route across central Europe to obtain the highly valued raw material.

Another case of the wide spread of items of value without any connection to a colonial history is the abundance of Arabic silver coins found in the medieval settlements of northwestern Russia and in Scandinavia. These regions were never colonized by Arabs. The coins were brought to northern Europe by Viking traders. The connection with the Arab-controlled markets was established along the River Volga, which was called the "route to the Arabs" by the Vikings. Trade goods are indeed a measure of exchange, but not forcefully an indicator of colonization. Trade relations emerge when there are articles to be exchanged, and there has to be a consensus about their value, among providers of such articles and those who are interested in obtaining them. The ancient Egyptians could extend their trade network under two conditions: (1) they possessed the resources and the know-how to offer goods of value for people outside Egypt, and (2) there existed cultures outside Egypt that were advanced enough to offer goods for exchange themselves. The reason for the wide spread of Egyptian trade

goods throughout the Mediterranean region is the presence of advanced civilizations that had quality goods to offer in exchange for Egyptian goods.

Among the prominent trading partners of Egypt in the second millennium BCE was Keftiu, which is the Egyptian name for ancient Crete. Keftiu is mentioned in Egyptian texts, which made Bernal jump to the conclusion that Minoan Crete must have been an Egyptian colony. Except for an abundance of Egyptian trade goods there are no signs of any colonization in Crete. The patterns of Egyptian colonialism outside the Egyptian heartland can be studied in the south, in Nubia, where Egyptian political supremacy was established in the mid-second millennium BCE. Visible markers of the former Egyptian presence in Nubia are, still today, the ruins of temples that were erected for the divinities of Egypt. Wherever an Egyptian community existed in antiquity, there was a temple.

An Egyptian temple was more than an architectural complex, it was a marker of Egyptian identity. The ancient Egyptians believed that, in the presence of a sanctuary for an Egyptian deity, it was possible for them to "breathe" the spirit of the Egyptian way of life, infused with a sense for righteousness and proper social conduct. The guardian of this process of unfolding Egyptian identity was the goddess Ma'at who is sometimes shown with a feather crown and sometimes with wings, allusions to the light medium "air" as the vehicle for the Egyptian spirit. In this view, an Egyptian temple was an anchor of spiritual as much as cultural identity. Given this foundational property of the Egyptian way of life, it is conclusive that archaeologists have unearthed the ruins of numerous temples in areas of a former Egyptian political presence outside the heartland. That is true for Nubia and also for certain areas in the Near East (e.g., in Palestine) where Egyptian political influence was felt as far north as Syria. However, there are no remainders of Egyptian temples neither in Crete nor in all of the regions with Greek settlements, the Greek *oecumene* (living space). This is a clear indicator that Egyptian communities were absent from the Greek *oecumene* and, thus, there was no basis for Egyptian colonial rule. The claim, made by Bernal (2006: 564), that Athens was allegedly founded as "the daughter city of Sais," the important trading port in the western area of the Nile Delta, lacks any archaeological evidence.

On the contrary, there is evidence for the presence of foreigners from the Aegean islands, from Crete in particular, in Egypt. They might have come as traders or artisans who offered their professional skills to Sesostris II (19th century BCE), king of the 12th dynasty. There is another clue to the understanding of what kind of relationship existed between Egypt and ancient Crete: Minoan pottery (especially Kamares ware) dating to the 19th to 17th centuries BCE.

Kamares ware would appear to have two major sources in Crete—the palaces at Knossos and Phaistos. However, the route and method of entry for this pottery into Egypt has given occasion for much discussion, and several

possible routes have been suggested. These include the journey from Crete to a Delta port and then on to Memphis or another centre by an inland river route; from Crete to Cyrene and thence along the North African coast to the Nile Delta; or from Crete to the Syrian coast, sometimes passing via Cyprus, and following the coasts of Anatolia and Palestine to the Nile Delta. (David 1986: 181)

There is something intriguing about the Minoan pottery in Egypt—it was obviously so fashionable that its style was imitated by Egyptian potters using local clay. This says something about the nature of the trade relations between Egypt and Minoan Crete as equal partners.

Key Technologies in the Greek World as Achievements without Egyptian Patronage

The rulers of Egypt waged many wars with their southern neighbors, the Nubians, over political supremacy in the region, and, eventually, they conquered Nubia and established colonial rule (contribution on the pyramid builders). The Egyptian presence in Nubia brought key technologies to the country, among them Egyptian-style architecture and ceramic ware. The Nubians later adopted the hieroglyphic script for writing their local language, Meroitic. The Egyptians never attempted to occupy Minoan or Greek territory, they never established colonial rule there, and, consequently, there is no evidence for the transfer of Egyptian technologies (i.e., architectural techniques and styles, writing) as the typical markers of foundational influence.

Architecture

Basic architectural forms were developed, independently, in Egypt and in the Minoan-Mycenaean world, and, later, the two regions experienced an exchange of inspiring influences moving in either direction. There is no echo of a one-way irradiation of Egyptian technical know-how into the Aegean. On the contrary, the idea has been entertained that perhaps the most typical Minoan type of architecture, the layout of a labyrinth, might have been imitated in Egypt. As David argued:

In architectural and artistic terms, the Middle Minoan and Middle Kingdom Periods saw a perhaps constant exchange of ideas and techniques. It has been suggested that some architectural connections existed during this period, and that these were perhaps evidenced in buildings such as the Labyrinth built by Amenemmes III [18th century BCE] at Hawara, which later visitors compared to the Labyrinth of Minos at Knossos. (David 1986: 182)

Writing Technology

The ancient Aegean scripts (i.e., Cretan hieroglyphics, Linear A, Linear B) and their derivations (i.e., Cypro-Minoan, Cypriot-Syllabic) find their inspiration in the much older tradition of the Danube script, and these scripts are unrelated to the writing systems that were in use in ancient Egypt (i.e., hieroglyphs, Hieratic, Demotic). The name "hieroglyphics" for one of the Cretan scripts refers to the use of picture symbols as signs of writing and to the religious function of this kind of writing, but the Cretan hieroglyphs are not historically related to the Egyptian type of hieroglyphics. The alphabetic script was not transferred to the Greek world from Egypt either, but this specialized technology was elaborated in the Near East and reached Crete with Phoenician trade goods in the 10th century BCE. The Greeks had known literacy before adopting the Phoenician script, which was a more precise tool for rendering the sound structure of language in comparison to the older Aegean forms of syllabic writing. The Greek alphabet emerged in the eighth century BCE, and, in the aftermath of the conquests of Alexander the Great in the fourth century BCE, this script became known in the Near East and in Egypt.

With the expansion of Greek culture and language during the Hellenistic era, the Greek script eventually was adopted to write the Egyptian language (in its late phase of development). The alphabet that was elaborated for the local language in Egypt at the time when Christianity spread in the Nile Valley is known as the Coptic script, with its signs derived from Greek letters. The name "Coptic" reflects the Greek name for the Egyptians, Aiguptioi. The Coptic alphabet was used in Egypt to write religious texts in the Christian tradition until the advent of Arabic language and alphabet in the seventh century CE. As a spoken language, Coptic declined in the late Middle Ages. As a sacred language of the Coptic church, it continues to be used for reading religious literature.

References and Further Reading

Anthony, David W. *The Horse, the Wheel and Language. How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton, NJ: Princeton University Press, 2007.

Bernal, Martin. *Black Athena: Afroasiatic Roots of Classical Civilization: The Archaeological and Documentary Evidence*, Vol. 2. Piscataway, NJ: Rutgers University Press, 1991.

Bernal, Martin. *Black Athena: Afroasiatic Roots of Classical Civilization: The Fabrication of Ancient Greece, 1785–1985*. Piscataway, NJ: Rutgers University Press, 1987.

Bernal, Martin. *Black Athena: Afroasiatic Roots of Classical Civilization: The Linguistic Evidence*, Vol. 3. Piscataway, NJ: Rutgers University Press, 2006.

Blok, Josine H. *The Early Amazons. Modern and Ancient Perspectives on a Persistent Myth*. New York: Brill, 1995.

Budja, Mihael. "The Process of Neolithisation in South-Eastern Europe: From Ceramic Female Figurines and Cereal Grains to Entoptics and Human Nuclear DNA Polymorphic Markers." *Documenta Praehistorica* 32 (2005): 53–72.

Burkert, Walter. *Greek Religion: Archaic and Classical*. Cambridge, MA: Harvard University Press, 1985.

Cavalli-Sforza, Luigi Luca, Paolo Menozzi, and Alberto Piazza. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press, 1994.

David, Rosalie. *The Pyramid Builders of Ancient Egypt. A Modern Investigation of Pharaoh's Workforce*. Boston: Routledge and Kegan Paul, 1986.

Dihle, Albrecht. *A History of Greek Literature from Homer to the Hellenistic Period*. New York: Routledge, 1994.

Gimbutas, Marija. *The Civilization of the Goddess. The World of Old Europe*. San Francisco: Harper, 1991.

Gimbutas, Marija. *The Goddesses and Gods of Old Europe, 6500–3500 B.C. Myths and Cult Images*. Berkeley: University of California Press, 1982.

Gimbutas, Marija. *The Language of the Goddess*. San Francisco: Harper, 1989.

Haarmann, Harald. *Early Civilization and Literacy in Europe. An Inquiry into Cultural Continuity in the Mediterranean World*. New York: Mouton de Gruyter, 1995.

Haarmann, Harald. "Indo-Europeanization—The Seven Dimensions in the Study of a Never-Ending Process." *Documenta Praehistorica* 34 (2007): 155–75.

Haarmann, Harald. "Writing from Old Europe to Ancient Crete—A Case of Cultural Continuity." *Journal of Indo-European Studies* 17 (1989): 251–77.

Haarmann, Harald. "Writing Technology in the Ancient Mediterranean and the Cyprian Connection." *Mediterranean Language Review* 9 (1997): 43–73.

Haarmann, Harald, and Joan Marler. *Introducing the Mythological Crescent. Ancient Beliefs and Imagery Connecting Eurasia with Anatolia*. Wiesbaden, Germany: Harrassowitz, 2008.

Jasanoff, Jay H., and Alan Nussbaum. "Word Games: The Linguistic Evidence in *Black Athena*." In M. Lefkowitz and G. MacLean Rogers, eds., *Black Athena Revisited* (pp. 177–205). Chapel Hill: University of North Carolina Press, 1996.

Karageorghis, V. "Notes on Some Mycenaean Survivals in Cyprus during the First Millennium B.C." *Kadmos* 1 (1962): 71–77.

Katičić, Radoslav. *Ancient Languages of the Balkans*, Vol.1. The Hague and Paris: Mouton, 1976.

Lambropoulos, Vassilis. *The Rise of Eurocentrism. Anatomy of Interpretation*. Princeton, NJ: Princeton University Press, 1993.

MacLean Rogers, Guy. "'Quo vadis?'" In M. Lefkowitz and G. MacLean Rogers, eds., *Black Athena Revisited* (pp. 447–53). Chapel Hill: University of North Carolina Press, 1996.

Marinatos, Nanno. *Minoan Religion. Ritual, Image, and Symbol*. Columbia: University of South Carolina Press, 1993.

Marler, Joan, ed. *The Danube Script. Neo-Eneolithic Writing in Southeastern Europe*. Sebastopol, CA: Institute of Archaeomythology, 2008.

Osborne, Robin. *Greece in the Making 1200–479 BC*. New York: Routledge, 1996.

Shearer, Ann. *Athene—Image and Energy*. New York: Viking Arkana, 1996.

5

China's head start in technological innovation was retarded by its efficient and centralized imperial government.

PRO	Talaat Shehata
CON	John Lee

PRO

To empathically state, as Jared Diamond alluded to in his heralded 1999 book *Guns, Germs, and Steel*, that China's early technological innovations were significantly retarded by an efficient and centralized government is to merely grasp one important factor in the entire argument put forth by Diamond. He clearly understood the multivaried reasons as to why China failed to attain that *take-off* moment, as later occurred in Europe. He also understood the important impact that the centralized and efficient imperial government, throughout a large swathe of China's history, had in China's dynamic society. Diamond states:

> [B]ecause of geographic factors—such as China's relatively smooth coastline, its lack of major peninsulas as large as Italy and Spain/Portugal, its lack of major islands as large as Britain and Ireland, and its parallel-flowing major rivers—China's geographic core was unified already in 221 BCE and has remained unified for most of the time since then, whereas geographically fragmented Europe has never been unified politically. (2005: 374)

He further elaborates that "that unity enabled China's rulers to command changes over a larger area than any European could ever command—both changes for the better, and changes for the worse, often in rapid alternation (hence, 'lurching'— the term is used in its neutral strict sense of 'swaying suddenly from side to side,' not in its pejorative sense of the gait of a drunk person)." Diamond concludes that "China's unity and decisions by emperors may contribute to explaining why China, at the time of Renaissance in Europe, developed the world's best and largest ships, sent fleets to India and Africa, and then dismantled those fleets and left overseas colonization to much smaller European states; and why China began, and then did not pursue, its own incipient industrial revolution" (Diamond 2005: 374). An important factor to always remember whenever one attempts to better understand early China's development, which should cast a much clearer light on

103

Yali

Jared Diamond starts his book *Guns, Germs, and Steel: The Fates of Human Societies* (1999) with mention of his talks with a Papua New Guinea politician Yali. A community leader, Yali was active in the cargo cult in the Madang area, being long recognized as the local leader of it. He was president of the council at the opening of the Local Government Council House in 1965 and contested the 1968 House of Assembly elections for the Rai Coast electorate, with his supporters viewing a victory by him as one of the signs of the cargo millennium—their disappointment at the results led to many local disturbances.

Diamond found that Yali, although he had never been to university nor ever left Papua New Guinea, had an insatiable desire to learn, posing the question on why there were such major differences between the lives of people in New Guinea compared with those in Europe or the Americas. This became the inspiration for Diamond's book.

---

modern China's recent and future development, is that the rulers had worked out a unique formula in creating stability and maintaining control over their large swathe of territory. Social control and moral regulations were able to very strongly bind the local powers within the country to the core or central administrative power of the state. A wide and expansive network was created, which had at its core a meticulously constructed social, moral, and ritual set of relations throughout the society. In times of crisis, which were regular, during China's rich and long history, these social, moral, and ritual relations kept China's elites and subjects powerfully bonded, even into the late days of their withering empire from the mid-18th, the entire 19th, and the early 20th centuries. It had more efficiently deeply saturated and penetrated the individual and family behaviors of all of China's diverse and ethnic subjects; even much more than the influence the county governments had in each of these regions and its ever-miniscule affairs, which in itself, was quite significant.

Therefore, the primary core and glue that held it all together for the Chinese, with much of the turmoil that often surrounded them and caused them to learn to adapt and coexist with numerous contradictions within the fabric of their society, was the population's consistent reliance on the efficient and centralized nature of their imperial government. That struggling and "lurching" quality to the dual nature of their national character ultimately caused them to gradually fall behind the increasing European expansion from the 16th to the late 20th centuries, in the political, economic, social, and scientific realms. What had evolved then with that retarded factor could also evolve into China's near and distant future in the 21st and 22nd centuries, especially with the all-inclusive role that the post-1949 communist state has sought to continue and

aggressively monopolize and propagate, especially 20 years after the crushing of liberal democratic dissent in Beijing at Tiananmen Square.

If one were to return to the earliest seeds of the Industrial Revolution in Europe, it would deductively and easily be surmised that the inventions of the time all came from China. Examples would be the compass, gunpowder, and the printing press. That being the case, the puzzling question to a few scholars in the area then was why didn't China, with its scientific innovations, not experience its take-off historical "moment" into the Industrial Revolution *prior* to Europe? Prior to Europe's take-off moment, China was the unrivaled global leader in science and technology. All these great accomplishments were perceived by the Chinese authorities as important yet uneventful. Europe, on the other hand, enamored by these new scientific developments and technologies, would gradually eclipse China, be completely transformed, and forced to take a new path in its history. The printing press allowed for more precise and error-free old tracts to be reprinted and disseminated on a much wider scale. It also made it much less expensive for the growing literate consumers to acquire a larger volume of written works than they had ever had access to, or for that matter, were permitted to own, in their earlier history. Then, with the advent of the Renaissance, a more classical humanist literary revolution would emerge, which in time led to the Protestant Reformation. With the advent of the Reformation, individual members of society were encouraged, if not implored, to base their ultimate religious salvation on their private readings of the Bible—a very transformative European historical moment, indeed. It would be followed by the Enlightenment period, then the age of revolution, with its sister 18th-century revolutions—the American Revolution of 1776 and the French Revolution of 1789. Earlier, the acquisition of gunpowder made castles' large fortress walls obsolete. Those who had access to these new explosive means could level any fortress within a few seconds. The aristocratic classes stood on very shaky grounds, to say the least. The winds of a new social order appeared to pick up speed in the distant new horizon, as the thick dense smoke above the demolished fortress walls began to slowly recede. Finally, the mastery of the compass would allow European powers to extend their individual and collective reach to other lands for their raw materials and expanding markets.

All that Europe was about to partake of the Chinese, as was often their manner, had kept a tight lid on for centuries. The potential of the compass was never much elaborated, nor was that of the use of gunpowder. As for the art and revolutionary nature of printing, despite it being downplayed by the Chinese authorities, it would, as in Europe prior to the third century, influence the Confucian and later the neo-Confucian classics. As would occur in Europe centuries later, printing in China brought to fruition new literary prose and writings, along with humanistic art forms, which had remained neglected for almost 1,000 years. Neo-Confucianism, as with Protestant Reformation and its primary attacks

against the Vatican and Catholicism, would have its drawn-out battle with earlier Confucian traditions and Buddhism. When the dust settled, neo-Confucianism, with the earlier teachings of Mencius (370 to 290 BCE), had found its voice. It completely individualized the faith, traditions, and practices of the ancestors and embraced Mencius's ideas that humans were born *tabula rasa* and, therefore, were basically good. For neo-Confucianists and most of the Song Dynasties inhabitants, it was a clear green light for anyone in society to mix her or his beliefs with the daily affairs of the community. Buddhist monks were not bestowed with any more heavenly qualities, as were their counterparts in the priesthood in the Catholic church centuries later, and they found their influence and power diminished within the Protestant Reformation. A "studious," monk-ish, single, and reflective life didn't accord any of them in their separate faiths and societies an *exceptional* standard by which the dedicated faithful needed to measure up to regularly in their daily lives. In life's daily struggles, they all shared in the common human bonds. Finally, the neo-Confucian system, with the advent of the print revolution and its regular dissemination of an eclectic inexpensive supply of information throughout China's urban and rural sectors, allowed for all highly motivated and qualified Chinese who resided in these widespread sectors to compete for prized bureaucratic positions.

But, on the whole, the imperial government of China persisted in the tradi-tional policy of constructing a silent wall, which surrounded many of these and other similar innovations. China had possessed most of these innovations during the Tang Dynasty (618–906), at least 600 to 800 years prior to Europe's acqui-sition of them, in the late 14th and early 15th centuries. But, with the arrival of the Song Dynasties (960–1279), China experienced the foundational changes that England would later experience in the late 18th and early 19th centuries, immediately prior to its economy's take-off phase in the earliest period of the Industrial Revolution. Even in the early 19th century, the ecological pressures that China faced around the Pearl and Yangtze rivers delta were not very differ-ent from those that other European countries found themselves contending with. The level of relative parity was quite evident in the areas of population growth, increased income, and overall consumption. But in these diverse historical peri-ods within China and Europe, one could also accurately conclude that they both underwent their *unique* technological and commercial revolutions. How it impacted each of them within their political, economic, ecological, social, and cultural realms, as the reader shall see, was quite different.

To best understand China's and Europe's separate historical evolutions, one needs to view them within a separate distinct process. As ridiculous as it may seem to a European and Western reader to view most, if not all, historical developments in Europe as being impacted by China, it is equally ludicrous to view most, if not all, historical developments in China as being primarily effected by Europe. Their early histories did not show a strong symbiotic

relationship. As a matter of fact, it was very weak. By 1000 CE in the northern and then shortly thereafter in the southern Song Dynasties, what is often attributed as one of the prime prerequisites in describing a society as undergoing the earliest stages of an Industrial Revolution, as later was witnessed in Great Britain in the early 1800s, the processes of mechanization and mass production thrived in China. There has been archaeological and historical evidence to support that argument. Over 16 million exactly shaped and produced arrowheads were discovered in an age-old arsenal manufacturing plant. Other innovations were also discovered in the northern region of the country. There were remnants of machines that were operated by the use of belt transmissions. These belts were often created by using a waterwheel contraption that created the finest of yarns by twisting large, rough, and thick volumes of hemp rope fibers. The technique used often included 32 identically structured spinning heads that rotated together in identical motion, as can be found in most early and late industrial textile factories in Britain and Europe. The important factor here is that instead of using a burdensome amount of human labor to create the yarn produced by these machines, it was reduced to the most basic of early industrialization's components (i.e., inanimate metal, steel, and wooden devices, with a human supervisor or two, to oversee the entire operation).

Also, during the Song Dynasties, the most sophisticated and technological sea-faring global fleets were in existence. By the beginning of the 16th century, with their centuries-old mastery of compass use, the Chinese merchant marine and naval ships had established total dominance over the north Pacific Ocean lanes and much of the Indian Ocean lanes, which circumnavigated the Indian subcontinent and reached as far as Africa's eastern shores. It was seamless. They didn't need a Vasco de Gama to circumnavigate the Cape of Good Hope to reach the East and many of the Indonesian islands, where Europe would clamor to acquire its fine silks and spices. The Chinese were already there. They were the "East" that all future West European navigators, including the Milanese Christopher Columbus, often stumbled over themselves to find new ways to get to their natural resources and finished goods. Most of the goods acquired by the Chinese in their eastern waters and lanes were much less expensive for the average Chinese merchant and mainland consumer than if they had traveled to Spanish, Portuguese, or Venetian ports to acquire the goods they needed. Why should they have bothered? They wisely didn't. The "fire-sales" were regularly taking place closer to their shores. All the imperial government needed to do for China to maintain its reach, as they did with well-known Ming Dynasty Admiral Zheng-He, was project Chinese naval might, as occurred in 1405, with gigantic displays of newly constructed armadas and ships as they slowly traveled out of the Yangzi River and into the East China Sea. The imperial government continued to regularly build, upgrade, and reinforce these naval and merchant marine vessels to make certain that the North Pacific region, the seemingly never-ending

The world's earliest surviving printed book is this version of the Buddhism scripture known as the *Diamond Sutra* produced with wood-block printing. The Tang Dynasty-era sutra consists of individual sheets of printed text and a frontispiece illustrating the Buddha surrounded by acolytes and disciples. (The British Library/StockphotoPro)

archipelagos and islands in the area, the Indian Ocean, and the East African coasts, remained stable and secure for continued profitable commercial acquisitions and gains for their nationals.

China's economy, therefore, during this early preindustrial period, remained strong until the mid-18th century. Throughout that period, Europe's economy was in constant flux, ever expanding and contracting. But it was never to be measured within the solid stable parameters within which China's economy sustained itself. It appears that among other important factors, the slow demise of China's economy vis-à-vis Europe began once China tied its finances to the silver trade. The Europeans, especially the Spanish, after their expansion into the New World, acquired obscene sums of silver from Mexico and Central America. Used as the common currency of trade on the American continent, Spanish and European powers, through their combined global explorations and conquests, began to trade with the most stable and dependable global economy at the time in much larger volumes—that of China. After the passage of close to 300 years, from the 1500s to the mid-1800s, Europe's economies had significantly developed beyond the backwater and periphery status it had occupied and resigned

itself to, for much of the past 1.5 millennia, in relation to that of China's solid core economy.

Throughout that period, especially during the early to mid-18th century, the silver trade grew, with China acquiring over 48 percent of the world's silver through its sale of its finished goods, such as silk, tea, and porcelain, to the Europeans who would then sell these goods at a significant profit to their subjects in Europe and throughout their conquered colonies. Eighty-five percent of the world's silver was mined and produced in Mexico, Peru, and Bolivia. This alone should be an excellent indication as to the dynamic nature of China's economy on a global scale. But, also, between 1527–1821, by allowing their fiscal system and economy to operate through the adoption of the silver standard, China's different imperial governments were able to double the currency's value in their economy and in their relations with the outside world. There was an attractive profit to be gained by local Chinese merchants through the exchange of gold in the country for the tons of silver arriving at their shores. This policy by China caused a pronounced surge in silver production in Central and Latin America and with its geographic neighbor Japan. Silver shipments to China's population began to increase. This was a consequence of the other more perishable trading that continued between China and Spain. The Spaniards were bringing to Chinese shores high-yielding crops, such as sweet potatoes, peanuts, regular white potatoes, and Native American corn. The Chinese population of 150 million in the mid-1700s, within the short span of 100 years, had tripled to 450 million by the mid-17th century.

Within this ongoing monetary and commercial linkage with global markets, from the early 16th to early 18th centuries, China's fortunes began to experience a serious setback by 1810. It was during that year that the world's silver supplies, as a result of the Napoleonic Wars and Latin American revolutions and independence movements, began to critically decline. By being so closely tied to the global silver market, China's economy nosedived into a depression. It would take almost another 200 years for it show any measure of recovery, during the Great Recession of 2007 that significantly afflicted the United States and other former European powers. By 1775, as mentioned earlier, Japan had stopped exporting silver. Silver from Burma and Annam were not at the requisite level, that the Chinese imperial government had grown accustomed to. Latin America, with the silver mines in Mexico, Bolivia, Peru, Chile, and Argentine, to mention a few, became China's only haven for its much needed silver. The silver that it found itself relying on was in the form of ingots. It also found itself having to depend on Mexican dollars, which kept a bimetallic system of currency functional. For payment purposes, this system required that all the mined silver be exchanged into copper cash. So, even though the imperial Qing government had total control over the coinage of copper, it was a factor that was completely out of their control, as were the merchants who established and sustained the price of silver in the market. This monopoly by these private

merchants over such an all-important metal and its market and exchange rate in the local, regional, and global economy set the pattern for the rate of exchange between silver and copper coinage. The imperial government found itself, against its deepest intentions and wishes, sitting idly by as China's silver supplies from Mexico, Bolivia, Argentina, and Chile began to significantly shrink. The tax burden on China's subjects and economy, as a result, increased. This caused prices for all goods to increase. As a direct consequence, the manufacturing sector significantly declined. This, of course, was cause for a significant spike in unemployment. To add insult to injury, established Chinese staples as silk, ceramics, porcelain, and tea, which were the engines of China's economy over the past two centuries, fell victim to a fickle global market. For a few brief decades, in the 19th century, from 1857 to 1886, China was able to turn its fortunes around by stabilizing its net trade balance (or what in hindsight appears as the imperial government's last effort to grab for straws to tread water and stay alive) with the doubling of the volume of opium being imported. But that should in no way be regarded, as has often been done by many scholars in Chinese economic history, as being the main contributor to China's economic collapse by the late 19th century. The failure in take-off of China's economy and the collapse of the state, as evidenced, had been unfolding since the early 18th century. Opium in the late 19th century became a very late symptom of the main problem in China's economy, which was maintaining its solid financial liquidity instead of *itself* being the main problem.

Having to tread and cope with these serious financial misfortunes, China's imperial government had other devastating problems to contend with, such as the 1810 climate changes and serious drought conditions, which significantly contributed to the famines of 1810, 1811, 1846, and 1849. It is estimated that close to 60 million Chinese perished within these four famines. With the Taiping Rebellion as a reaction to these catastrophic events, and most importantly, the fast shrinking clout of the silver trade affecting China's daily economic affairs, the decades of the 1850s, 1860s, and early first three years of the 1870s proved to be equally cataclysmic for China and its imperial government. It is estimated during those three decades of turmoil that China lost close to yet another 60 million Chinese. This should also factor in Japan's growing strength after the modernization it had undergone during the Meiji restoration. Along with other European powers, it was beginning to pose a threat to China's hegemony in her traditional outposts in the Pacific. But the final straw for the imperial government, which guaranteed them and their society a free-fall that lasted until the Communist Revolution in 1949, was when the seriously tested and stressed imperial government decided to "momentarily" shift its economy from its usual state directives regarding grain's direct shipment to many of its starving populace and instead allowed a few of the monetary charities to assume that responsibility. This guaranteed the complete collapse of the system. Chaos ensued, with

vast numbers of the population setting out to debark the forests, stripping houses and fields along their raging and desperate paths in order to find something to eat. All that ensued as a direct result of the grains being distributed by charities running out way too soon and the imperial government's gross negligence and mind-boggling mismanagement. To put it all in context, it was the late 19th-century's Chinese version of the former Bush administration's pathetic and tragic handling of the Hurricane Katrina disaster. For the next 70 years, it would guarantee that no take-off historical moment would ever occur for China's economy, as it had earlier for Britain, or any of the European powers. The imperial government became too overwhelmed with the myriad factors just referenced, which as a whole created that perfect storm from which they would find it totally impossible, over the immediate short term, to recover from. As a direct result of that perfect storm, Diamond's "efficient and centralized Imperial government" ceased to function. Diamond's view, therefore, should be viewed in that light. That is, if that take-off historical moment failed, it was not so much that the imperial government, over the century, had lacked efficiency or was even centralized. His point is that the imperial government had instilled in the Chinese population and different ethnic groups the strongly held and secure idea that they always had the imperial government to reach out to for all their needs, especially during periods of crisis and peril. But that rule, as the centuries unfolded, always had a double-edged side to it. When it worked, it was vibrant, refreshing, and challenging to the average Chinese subject. But when it didn't, the unraveling process was simply excruciating and often catastrophic and cataclysmic—as witnessed from the early 18th century, through the 19th and 20th centuries.

As Diamond would have his readers understand, it was ultimately that total lack of efficiency and centralization by the imperial government, which most Chinese were not usually accustomed to, that created that very deep sense of dissonance. With the failure of the imperial government, order under the heavens, as the average Chinese elite and subject knew it, ceased completely to exist. With the last imperial government's failure in the late 19th century, it had not allowed for any political, economic, or social safety net of which the majority of the population could make solid use. Instead, it created the most nightmarish scenario that most Chinese women and men have ever experienced throughout most of their richly shared and textured history. And that is an abysmal vacuum that would continue to inflict further agony, pain, and tragedy on China for much of the next century and a half, even after Chairman Mao Tse-Tung's Marxist-Leninist Revolution in 1949. China would have to wait out those long grueling years until the first glimpse of an opportunity presented itself, with the Great Recession of 2007.

With modern China's growing financial and economic role, once again, as a direct result of the ongoing woes experienced by the United States and Western Europe since December 2007, many shrewd economic analysts believe that

China will overtake the United States as the sole economic superpower by the year 2027. As has been shown in its rich and textured history, if China's "imperial" elites and government can't find a way to stop functioning according to Diamond's eloquent "lurching" description, then the past might sadly prove to be a prologue. Modern China and its imperial government remains uncertain as to whether it should proceed forcefully as a capitalist economic superpower or rein in its successful capitalist "urges," with its political and often repressive Marxist-Leninist state. The state has lurched from one moment, allowing its modern citizenry more freedoms in their domestic and international economic activities, but, in the same breath, crushing any form of liberal and democratic political forms of dissent, as it so successfully did at Tiananmen Square in 1989. Adopting a consistent policy of ideological intransigence, the Chinese imperial Marxist state used that form of intimidation to maintain control of its domestic realities. This policy has been regularly implemented to also keep at bay any pressures brought to bear on them by any foreign powers. As a result, the elites in the state have convinced themselves that both domestic and foreign policies have grown into a seamless unit as they've continued to lurch along. At an annual growth rate of 7 to 8 percent a year, and retail sales rising in the first quarter of 2009 to 15 percent, China's imperial state has seen no reason to seriously reconsider its domestic and foreign policies. Quite aware that they hold at least $2.3 trillion of U.S. debt, it's not unusual to hear statements such as that of the governor of the People's Bank of China, that the International Monetary Fund might one day reconsider the displacement of the U.S. dollar. The primary reason for such a statement was that Chinese Premier Wen Jiabao had expressed his and fellow imperial government officials' concerns that China's vast reserves were held as dollars. They had hoped that by issuing such a public statement, the U.S. government would issue more TIPS (Treasury Inflation Protected Securities). These TIPS would perform a different function than most conventional bonds, in that in the event of a significant spike in inflation, the owners of the bonds would be shielded. Despite their knowledge of the bigger role others are expecting them to play and that they themselves find themselves playing, they could not imagine, at the present time at least, ever being able to throw overboard the U.S. dollar or replace the dominant U.S. role in the military or economic global context. Russia sought in the first BRIC (Brazil, Russia, India, China) summit meeting to have these four fast-growing economies reduce the world's and each of these country's collective reliance on U.S. government bonds. The idea floated by the Russian delegation was that the Chinese yuan would replace the U.S. dollar as the primary reserve currency. But, wisely, and possibly a subterfuge to buy more time, an official at the China Institute of Contemporary International Relations in Beijing summarized China's reticence on that very sensitive matter. He was quoted as saying, "American economic dominance has been badly wounded, but its economic strength and overall national

strength will remain far in the lead for the next 10 to 20 years" (Buckley and Rabinovitch, 2009).

With that thought in mind, China's imperial government has continued to lurch. With the global financial turmoil and drama playing itself out, China has allowed numerous loans-for-oil deals and endless large currency swaps. One currency swap was with South Korea. By the end of March 2009, for a total of 650 billion yuan or $95 billion, it had swapped currencies with Argentina, Indonesia, and five other nations. China also gave $45 billion in credit as loans-for-oil deals to Brazil, Venezuela, Angola, and Russia. This guaranteed China long-term crude supplies. With the former Chinese diplomat in Washington, Zhen Bingxi, writing in the *Chinese Journal, International Studies*, stated that "China can now apply its (foreign) reserves while Western economies are weak and helpless to extend aid to other countries in dire straits and thereby win partners"; political sway by China's imperial government through the use of its reserves in trade matters seems to be the order of the day. But, with a more balanced participation in the international monetary market and a realistically valued yuan, China has come to the gradual realization that it must play a much more forceful role (Buckley and Rabinovitch, 2009).

Despite such positive monetary gains for China's imperial government, it found itself constricted by its lurching habits. Recently the modern imperial government has found its sound economic policies undermined by its self-destructive political actions. It announced that all of China's new personal computers sold in China had to have software preinstalled that would restrict access to Internet sites that the government deemed "malicious" or "undesirable." If it is successful in such a policy, it could undermine many of China's hard-fought economic gains over the past three decades and make it totally uncompetitive with the other advanced global economic powers. Then, a few short weeks later, it lurched back to a policy that made the installation of such software and filters "optional." Then, continuing to add insult to injury, China's government, during the public's commemoration of the 20th anniversary of the Tiananmen protests, had its censors block any of its citizens access to Microsoft's Live.com, Flickr, Hotmail, or Twitter sites. With such continued old ingrained habits of lurching, China's current imperial government might find its lack of fit between its economic and political policies, in the long term, undermining all the hard work that its dedicated citizenry and the leadership had put into that long march forward, after the collapse of the last imperial government at the end of the 19th century. If modern China is unable to have its political liberties catch up with its economic liberties, then, in time, its citizens, and the world, might be witness to yet another catastrophic collapse from which the Chinese state will not be easily able to recover from in a much fiercer and more competitive postmodern world. In such a very probable set of events, it won't be only the Chinese, through their incessant lurching, that

would lose a very rich history and heritage that could have been shared with the rest of the world as a revived and dynamic Middle Kingdom, vigorously projecting it to future generations. Just as importantly, it would become the rest of humanity's tragic loss. For those precise reasons, the modern Chinese imperial government needs to step back, take a serious and conscientious inventory of its domestic policies, then take a deep breath, gain a strong grip, and permanently stop the lurching.

References and Further Reading

Bol, Peter K. *Neo-Confucianism in History*. Cambridge: Harvard University Press, 2008.

Bozorgnia, S. M. H. *Role of Precious Metals in European Economic Development: From Roman Times to the Eve of the Industrial Revolution*, Vol. 192 Westport, CT: Greenwood, 1998.

Buckley, Chris, and Simon Rabinovitch. "China Lays Road to Global Role with Economic Cement." Reuters Web site, June 11, 2009. http://www.reuters.com/article/idUSTRE55A0U320090611 (accessed May 25, 2010).

Davis, Mike. *Late Victorian Holocausts: El Nino Famines and the Making of the Third World*. London: Verso, 2002.

Diamond, Jared. *Guns, Germs, and Steel*. New York: Norton, 1999.

Diamond, Jared. *Collapse: How Societies Choose to Fail or Succeed*. New York: Penguin, 2005.

Freedman, Paul. *Out of the East: Spices and the Medieval Imagination*. New Haven, CT: Yale University Press, 2008.

Halliday, Jon. *A Political History of Japanese Capitalism*. New York: Pantheon, 1975.

Josephson, Matthew. *The Robber Barons: The Great American Capitalists 1861–1901*. New York: Harcourt Brace, 1934.

Kuhn, Dieter. *The Age of Confucian Rule: The Song Transformation of China*. Cambridge, MA: Harvard University Press, 2009.

Lewis, Mark Edward. *China between Empires: The Northern and Southern Dynasties*. Cambridge, MA: Harvard University Press, 2009.

Miles, Steven B. *The Sea of Learning: Mobility and Identity in Nineteenth-Century Guangzhou*. Cambridge, MA: Harvard University Press, 2006.

Mohide, Thomas. *The International Silver Trade*. Cambridge, UK: Woodhead Publishing, 1992.

Sen, Amartya. *Poverty and Famines: An Essay on Entitlements and Deprivation*. Oxford: Clarendon, 1982.

Tuan, Hoang Anh. *Silk for Silver: Dutch-Vietnamese Relations, Tonkin 1637–1700*. Leiden: Brill Academic Publishers, 2007.

Wilkins, Mira. *A History of Foreign Investment in the United States to 1914.* Cambridge, MA: Harvard University Press, 1989.

CON

With the Olympics recently completed in Beijing, China, this section of the chapter is appropriate for explaining China's recent growth and push into the world's consciousness. As Europe took advantage of resources, social structure, and trade relations, so too modern China has taken advantage of these same elements to pull ahead of world competition. The period of Great Disparity exposes the roots of European advancement and the causes of Chinese backwardness during the 16th century. The section will explore the history of this disparate growth and progress between the two continents.

In the process of world history, the great nations of the Eurasian continent grew into large empires; while on the European continent they continued to exist as nation–states in competition with one another. Within the past 500 years, these European states would outstrip and dominate the large empires of China and Japan through advanced technology, resources, and political systems. China had advance between 1000–1600 into an empire that was technologically advanced, culturally superior, and politically more organized than the European nations. However, the Europeans outcompeted them in all three areas between 1600–2000. Jared Diamond, a historian and writer of *Guns, Germs, and Steel*, proposed his argument that it was due to geographic factors that China stagnated and Europe progressed. It was an interaction of opportunity and necessity due to the environment. Eurasia's large land spaces allowed for the domestication of many more animal species than other continents (Diamond 2005: 5). Diamond explains that the east-west orientation of land, the availability of animal species, and the climatic similarities of the continent allowed for the growth of Eurasian civilizations (Diamond 2005: 6). He states that the large plains and vast territories that created the large empires and ponderous bureaucracies of China allowed them to stop advances because they had no external pressures to force them to adopt these advances (Diamond 2005: 432). Mountains, rivers, and oceans that forced countries to compete with one another, however, divided up Europe, and they could not survive as nation–states without keeping ahead of the other nations in technology, culture, and political system (Diamond 2005: 430). In addition, Europe took advantage of the availability of resources by creating divisions of labor—merchant class, craftsman, and a specialized writing class—that allowed the economic revolution to occur in Europe. My position is that Diamond ignores other cultural, political, and geographic elements that

---

Jared Diamond: Historian and Much More

Jared Diamond was born in 1937 in Boston, from a Polish Jewish family, the son of a pediatric hematologist. He initially planned to follow his father into medicine and completed his doctoral thesis on physiology. Then Diamond became interested in ornithology, which in turn led him to visit Papua New Guinea, writing several books on birdlife there. It was on one of these trips that he came up with some of the ideas behind his book *Guns, Germs, and Steel*. He is now professor of geography and physiology at the University of California, Los Angeles.

The advantage of Diamond's varied background was that he has been able to write about natural history, science, and modern history with ease. As well as English, he is able to read Finnish, French, German, Greek, Indonesian, Italian, Latin, Russian, and Spanish, as well as neo-Melanesian and Fore, spoken in the Eastern Highlands of Papua New Guinea.

---

slow China's advancement and help to increase Europe's advance into the modern era.

The Historical Background: The Renaissance of Chinese Inventions

The important question is why did China lag behind during the 16th and 17th centuries? It had developed into an advanced civilization with the development of much progress in the scientific and technological fields. In science, it had created inventions such as gunpowder, iron smelting, the spinning wheel, the plow, and the harness before the Europeans (Temple 1986: 215). The Chinese also had created silk spinning, cotton weaving, and other textiles (Temple 1986: 94). Furthermore, the Chinese had invented paper, paper money, books, and other forms of written communications (Temple 1986: 81, 99, 116). In mathematics, the Chinese had developed the decimal system (1400 BCE); invented the concept of zero that helped with computations; invented the idea of negative numbers, square roots, and decimal fractions (100 BCE); algebra (300 BCE); a refined value of pi; and the inventions of Pascal's triangle (1100 CE) (Temple 1986: 139–46). In physics, they developed the first compass dial (300 CE), magnetic induction (1100 CE), as well as the first law of motion (400 CE) before Isaac Newton had (Temple 1986: 156–70). Cheng Heng, astronomer royal, created the seismograph in 132 CE to detect earthquakes in China. In terms of transportation, China created the kite (380 BCE), manned flight using kites (283 CE) by Ko Hung, a taoist, the parachute created by Emperor Shun (1192 CE), paddle wheeled boat (500 CE), and canal building. In astronomy, the Chinese discovered sunspots, cartography (map making), solar wind, and astronomical instruments (Temple 1986: 29–30). In engineering, they created spouting bowls, belt drive, waterpower, chain drive, and salvage procedures. In domestic technology, they invented the fishing

reel, porcelain, matches, chess, and permanent lamps. In medicine, they discovered circulation of the blood, circadian rhythms, endocrinology, thyroid hormone, and immunology (Temple 1986: 123–35). Musically, they created the tuned bell, tuned drums, and timber. They discovered warfare techniques of poison gas, tear gas, crossbow, flamethrower, land mines, grenades, guns, motors, and repeating guns. These inventions show that more than half of the modern inventions derived from Chinese ingenuity, but this seems to have been forgotten more than several centuries later. The modern world was a combination of Chinese ingenuity and European improvements on these inventions.

Yet somehow, the Chinese lost their edge in the 14th and 15th centuries when the Europeans pulled ahead with the advent of the agricultural and scientific revolution where the Europeans took these inventions and applied them in ways that would lead to the Industrial Revolution of the 17th century. They took the inventions of the steam engine, the loom, and the agricultural techniques to form the foundation of industry that produced textiles and cotton in the factories of England and Germany. The foundation of this advancement was the European creation of objective science and capitalism that provided the framework for the interchange of ideas and the necessary investigation of natural phenomena. Objective science allowed the Renaissance scientists to create anatomy, chemistry, and other mechanical inventions. Capitalism created the merchant class that allowed the interchange of goods to raise the standard of living as well as allowed the competition of ideas between nations. Accordingly, the Europeans developed maritime inventions such as the compass and the astrolabe, which allowed for the international trade and imperialism of European ideas. As a result, the Europeans pulled ahead of China, which was struggling in the throes of corruption and a stagnant nation.

Critique of Diamond's Eurocentric Perspective

Many historians have created general arguments against the overall thesis of Diamond's book. The first argument criticizes the Euro-centrism of the book, which means it places too much emphasis on the dominance of European geographic factors. The historians criticized that he adapted too many of the theories from other historians and that the thesis is not original. Several historians have complimented Diamond's thesis, but attack specific arguments that Diamond proposes. The thesis that Eurasia dominated the scene because of its geography was challenged by Andrew Walton, who stated that Diamond refused to acknowledge that the population was very large compared to other continents, so by chance it would come to dominate the world (Walton 2006: 6). Walton also noted that it is not clear whether it was technological superiority or disease that allowed the Dutch and English to conquer Asia (Walton 2006: 6). It might have been diplomacy and cultural integration during the 19th century

as well as Chinese corruption. In addition, Diamond fails to account for the current studies in colonialism, imperialism, and world systems, which explain the differences in progress between the Chinese and Europeans.

It might have been diplomacy during the 19th century as well as Chinese corruption. James M. Blaut wrote a critique of Euro-centric historians called *Eight Eurocentric Historians* that discusses the weaknesses of famous historians such as Max Weber, Eric Jones, and David Landes, but he especially criticizes Diamond's book for unsupported generalizations (Blaut 1999: 301–99). One generalization is that the east-west axis theory as proposed by Diamond is wrong because there was mobility, and the distance is the same between the continents. The mobility allowed for trade communication necessary for competition. He also dismisses Diamond's idea of domestication of animals because only some animals were domesticated in the centuries later. These arguments were just attempts to denounce Diamond for laying too much emphasis on Europe as a causal factor for the advancement of Europe in the world.

The Political and Social Criticisms of Diamond

The political arguments are based on the exceptions to the rule about how geography created political systems. Certain exceptions exist, such as Africa and Latin America, that were subdivided into many different countries but did not develop the competitive governments and the capitalistic systems of Europe. As with all general theories, these examples do not fit the rule. Many different countries had different states and nations that competed within the continent. For example, Africa was divided into many states and so was Latin America, but they did not formulate or create the culture necessary for technological advancements (McNeill 1963: 363). Both of them had rivers and coal supplies that could provide the basis for an Industrial Revolution, but they did not exploit these resources. Other historians such as Victor Davis Harrison agree with Diamond to a point, but they state that Diamond ignores factors such as capitalism, individualism, rational debate, and political freedom as the seeds for the advancement of the European Industrial Revolution.

In addition to these particular theses, historians have criticized Diamond's lack of concern about internal events, trade, and other factors in his historical analysis. William H. McNeill wrote in *The Rise of the West: A History of the Human Community*, a book that describes the rise and influence of Europe in the past 500 years using a cultural, artistic, and social approach, how Europe contributed to the Westernizing of the world. He was criticized for being Euro-centric too. McNeill argues that Diamond fails to account for the Chinese population's reaction toward opium, which gave rise to peasant rebellions and finally to the Taiping Rebellion that sought to undermine the authority of the imperial court (McNeill 1963: 363–65). This threat affected the court's outlook

on foreign intrusion as well as adopting foreign inventions of the British. It also lowered the economic prosperity of the court that did not allow the Ming Dynasty to explore countries or maintain ties with the British. McNeill also explains that interconnections between countries were important in the backwardness of the Chinese. The Chinese wanted to maintain the image that the imperial court did not want to connect with other countries because China was morally, politically, and culturally superior over the Europeans (McNeill 1963: 363–65). These were cultural factors that influenced the fear and isolationism of the Chinese, which created an unwillingness to accept other inventions from the Europeans. Whereas the Europeans had no pretense, they would accept inventions and adapt them to their own usefulness.

R. B. Wong emphasizes the social structure and the convergence of capitalism as well as technology in his *China Transformed: Historical Change and the Limits of European Experience*. He said that too many historians try to make European development a norm and China's development a deviation from the norm. In an objective approach, he discusses state formation and economic development (Wong 1997: 265). He believes that the economies of 17th-century Europe and China did not differ very much until the Ming Dynasty. Wong also attempts to analyze the reactions of the various states to the problems of internal rebellion and intrusion of the foreigners. Wong says that the Chinese stratified social structure and prevented the adoption of capitalism as well as technology (Wong 1997: 265). The unity of Chinese society was not flexible enough to link the capitalist economy and the technological development. It also did not emphasize the individualism that was needed to create the competition and markets needed for industry. Thus the social structure and Chinese unity prohibited the expanse of capitalism necessary for the advancement of the society.

Kenneth Pomeranz wrote in his book *Great Divergence* another theory for the economic causes of the disparity. Pomeranz claims that the Europeans and Chinese had similar ecological situations that withdrew the geographic explanations for the divergence. He also emphasized that the economic trade was important for the European expansion, allowing them to explore markets (Pomeranz 2000: 207). He also states that Europe was able to use coal deposits that allowed the Europeans to sustain an industrial economy. It allowed the Europeans to put more effort into the energy-intensive industry. Furthermore, Europe had an input of trade and products from the Americas that helped to provide prosperity for the Europeans. As a result, the Europeans could use less labor-intensive economies than the Chinese because of these resources. The Chinese, however, were stopped by several factors. The outer reaches of Chinese land had an economic boom in the 1750s, but they could not transport this boom to the central Chinese economy, which stagnated and could not expand. They had no access to trade products like the Europeans. As a result, Pomeranz uses an economic and network argument that defies Diamond's geographic explanation for the divergence.

The Cultural Criticism of Diamond's Thesis

Diamond's cultural arguments are based on the theory that the overall Chinese culture contributed greatly to the divergence technologically. The Chinese emphasis on neo-Confucian thought's tradition, scholarship, and social position created a country that would not accept new ideas or create new ideas from the people.

The Confucian ideal of a man was based on this man's obedience to others and to the authoritative structure of China, especially in the ideal of filial piety. China was able to create a harmonious society that emphasized the group and connected the individual with the state. Obedience, harmony, and respect allowed the Chinese to create the vast society that slowed down the dissemination of ideas. With this authoritarian structure and emphasis on tradition, the need for new ideas was squashed at its inception (Needham 2004: 84). In contrast, the European emphasis on intellectual advancement and reason allowed for the advancement and acceptance of technology and creation of inventions. Capitalism was a European invention that vitalized European discovery and invention, while feudalism slowed China down. In another vein, the philosophical viewpoints of both countries differed in their view on the self and object relationship. Taoism viewed the subject or person as part of the universe or the objective world (Needham 2004: 226). They were united as one, and the person was subsumed into the greater whole of the culture and society as well as with nature. This viewpoint retarded the acceptance of new inventions as means to dominate nature and produce things to advance the Chinese society. The manipulation of nature was considered antagonistic to the harmony of man and the world, so the Chinese did not exploit the new inventions.

European thought was based on the separation of man from nature and the domination of nature as part of the Christian heritage. This division allowed the Europeans to analyze, manipulate, and innovate tools to work elements out of nature instead of working in harmony with nature. Most environmentalists will agree that the Christian concept of Adam's rule of nature and tending the Garden of Eden was a permission to exploit and dominate the elements and nature. As a result, the domination of Protestant thought and the revolutionary worldview of Christianity allowed for the growth of inventions and technological advancements in the world. Protestantism emphasized the exclusion of other gods as well as an individual approach to religion that allowed for the freedom of thought needed for science (Needham 2000: 224). However, because of the freedom of thought, European science lead to viewing things outside of nature, namely laws and principles. The Chinese never thought about anything outside of nature. In addition, the atomism of the Greek divided the spiritual and material world into particles and elements. However, the Chinese always thought of matter as a continuum with the philosophical theories of yin and yang and the "five elements." This conflict also is seen in the Western approach to medicine

and anatomy, where the body is classified and divided, and the spirit is separate from matter (Needham 2000: 81–82). However, the Chinese regarded the body as a microcosm of nature and a container of fluid energy. As a result of these attitudes, the climate necessary for the rise of science was never created. These ideas were squashed by Diamond because he thought they were proximate causes of Chinese cultural backwardness. In some sense, Diamond exhibits this emphasis toward science as an explanation of all things.

Another argument is that Diamond does such a short analysis of the elements that contributed to the major changes in modern history that he does a disservice to other historians, who have studied them in greater details. There is a broad criticism that takes this into account that says he dismisses the centuries of work of previous historians in imperialism, society, and the interaction of world histories with a few broad-brush strokes. He believed that the bureaucratic structure hindered the development of science during the Ming Dynasty. However, some historians have noted that previous dynasties had similar bureaucratic structures that foster scientific investigation even when looking to feed the poor peasants of China. Also, another criticism is that he forgets to account for the impact of minor events in their effect on world history. In line with this criticism is the fact that the Ming Dynasty stopped exploration of the world when a lightning ball invaded the emperor's Forbidden City, which the emperor interpreted to be a sign that he was losing the Mandate of Heaven in his projects. So superstition stopped the bureaucracy from exploring and developing their naval power and communicating with the outside through Tseng He's fleets. As a result of the lightning ball, China became isolationistic and rejected all exploration of the ocean. This isolationistic tendency was carried over to future dynasties that disallowed conquest and the introduction of new ideas into China. Other minor details were the fact that rebels who invaded the territory of the Chinese during the Ming Dynasty overwhelmed the Chinese state. It stopped inventiveness and created an insecure social climate. In other areas, especially proximate factors, Diamond does not acknowledge the immediate economic factors of the demise of the Ming Dynasty. The Ming had an insufficiency of silver because of the Silk Road trade, which led to inflation and destroyed the economy before the British arrived (Gernet 1982: 533). Furthermore, the Ming had to deal with piracy and barbarians during this period, which further increased the demise of the Ming Dynasty.

There are many historians who believe that the Europeans could not advance without the contribution of the Chinese. The Chinese gave the West their technological inventions, but did not see the great possibilities in them, nor were the conditions correct for the use of these inventions. Before the 15th century, Europe was a backward continent that did not have technological competence. There are several inventions that helped to create the Industrial Revolution of Europe in the 17th century that China shared with Europe: the plow, the horse harness, and the multiple seed drill, which made agriculture and farming more

efficient and faster. These three inventions helped to revolutionize European agriculture and increase crop output. In some history books, these inventions also helped to create the agricultural revolution that laid the foundation for the Industrial Revolution, because it improved health conditions, food production, and the quality of life that allowed the Industrial Revolution to take place in Europe. In other words, China gave Europe the seeds that would later allow Europe to advance and surpass the Chinese in inventions and technology. China also assisted in the development of modern chemistry through its advance studies in alchemy. China gave the Europeans ideas that would be expanded into the New World. In addition, the Chinese influenced the Enlightenment with their philosophy and political ideas. Voltaire and other philosophers admired Confucius and his ideals of the political state. They also admired and adopted the Chinese idea of civil service into the democratic government that took shape in the Enlightenment. Because of European expansion into the world and the use of smallpox-infested blankets as weapons, they were able to exploit the world for their own ends. In other words, the Chinese gave us the foundations and the roots to our modern civilization in the cultural exchange that was going on in the 16th and 17th centuries. Some historians believe that the Europeans stole Chinese advancements and used germs to advance their civilizations in imperialistic ventures around the world.

A Special Critique: The Wall Mentality Critique of Diamond

The Chinese built the Great Wall as a result of repeated attacks across its frontier, and Europeans did not have a history with a wall like the Chinese. This theory was adopted from Julia Lovell's *The Great Wall: China against the World, 1,000 BC–AD 2000*. The history was that the Mongolians attacked China many times during the advancement of Chinese civilization, which provoked China to build an indestructible structure to defend the country. However, the wall became a symbol of China's defensive and offensive attitude against foreigners. The wall became a symbol of Chinese imperialism—when they crossed the wall to attack the Hsing-nu and the wall became an outpost of invading armies. The wall also became a symbol of fear of outside invaders as well a symbol of strength. It represented the Chinese fear of the outside world coming into invade the unifying culture of the Chinese people as well as geographic invasion (Lovell 2006: 6–10). The forces threatened to destroy the structure of the Chinese society and Chinese stability, so the Chinese undertook reactions that suppressed new ideas from these foreign forces. Even more serious is the fact that China never actually achieved unity in the nation because of the incursions of the Mongolians and the Chinese reaction to them. The Chinese imperial government wanted to stop the spread of ideas with political suppression as well as military defensive measure. As a result, the wall became symbolic of the attitude that

Considered one of the world's greatest manmade structures, the Great Wall of China extends about 4,500 miles, from Shanghai Pass to Chia-yu Pass. Large parts of the wall date from the seventh through the fourth centuries BCE, and much of the wall was built during the Ming Dynasty. On July 7, 2007, China's Great Wall was added to the list of New Seven Wonders of the World. (Corel)

would exclude Western science and technology as well as pull China from advancing into the modern era.

The wall was built during the reign of Huang-Di, the Yellow Emperor of the Chin Dynasty. In an attempt to control the incursion of the steppe people of Mongolia or to expand his territorial power, the Yellow Emperor built the wall with the conscripted labor of his people (Lovell 2006: 47). He utilized the various resources around the countryside and used hundreds of thousands of his people to build the brick walls. Furthermore, he garrisoned the walls with thousands of soldiers who would protect the wall from invading Mongolian invaders. Some say it was to expand his empire into the northern reaches of China. As a result of this project, Huang-Di demanded taxes and money from his people, which in turn created a centralized nation to support the demands of building the wall. In the end, the empire broke into many pieces because of inheritance issues. The Han Dynasty continued the relationship with the wall by building it as far as Chinese Turkistan as a protection against the Hsing-nu. They built outpost and signal towers to defend the western borders against their opponent. However, they also attacked the Hsing-nu, which resulted in a defeat

at Longmenan and required them to send tribute to the steppe people in the form of princesses, booty, and riches. As a result, the wall became the pride of Chinese nationalism, but it also was a burden of unbearable sacrifice in men, money, and emotional security (Lovell 2006: 66).

During the Wei and Yan Dynasties, leaders, who were descendant from the steppe people, dominated the Chinese. Also, they created walls to protect their coming to power in Luoyang and they adopted Chinese culture. Later in the Sui Dynasty, the emperor exacted a tribute system that suppressed the intrusions of the people of the steppes. Then in their building projects, the Sui attempted to build more walls in their capital cities and also built canals that would help with transportation and connecting the vast empire. Later the men and the money that they put into the wall caused the collapse of the empire. The next dynasty, the Tang, decided not to build walls and they flourished in a Chinese renaissance that opened up China to the world. Persian horses, polo, and various religions flourished in the newly created empire. Even the wall creates a new genre of poetry known as the Western frontier poetry, which expressed the loneliness of the outposts around the wall as well as the cost of building the wall in human lives (Lovell 2006: 150–57). However, the rebels fought against the Ming, which would close the walls again during the Jin Dynasty.

Then the Chinese had their worst nightmare come true, which was the invasion of the Mongolians from the steppes of Mongolia in 1200. This is the invasion of Genghis Khan that overthrew all of the defenses that were put up by the Jin. The invasion is due to three factors. First, the Mongol tended to pillage and destroy entire populations of Chinese because they wanted the booty, not the people, to sustain their lives. Second, the allies that manned the walls decided to desert the defenses. Third, the Jin defenders had a false sense of security because of the walls (Lovell 2006: 159). Kublai Khan, however, fell to the temptation of Chinese culture and adopted the richness of China. Finally, the rebellions of the Red Turbans overthrew the Mongolians.

During the Ming Dynasty, the Chinese took pride in their country by proclaiming that China was the Middle Kingdom, the center of the world. Emperor Yongle built the capital Beiping in the north to observe and take care of the Mongolian threat. He built the Forbidden City, Imperial Palace, and made walls to isolate the emperor from his people. Then he connected all the walls on the northern borders once again using Chinese labor to build this wall. Between 1547–1550, Altan, a Mongol, attacked Peking, Shanxi, and Hebei, from which there were many casualties among the Chinese. The emperor reacted by building 1,200 towers and nine border garrisons, adding 6,000 kilometers to the immense structure (Lovell 2006: 262). The purpose of this wall was to identify, surround, and ostracize foreigners. The social and economic consequences were large for the Ming Dynasty. The Ming Dynasty fell, once again because of the fear of foreigners.

The Qing Dynasty inherited the same attitudes as their predecessor, that is that they were the rulers of all lands and they were inheritors of Genghis Khan's

empire. Their emperorship dominated and they did not accept the small island of Europeans that came to their empire to lift trade restrictions in 1793. They still refused when European gunboats traveled up their rivers, because they felt that China had military superiority, and they persuaded the barbarians to accept their civilization as well as their tributary system. However, they forced their country to into isolation from Western imperialism, because they felt they had no need for Western science. As a result the wall mentality stopped the advancement of Chinese science. The Chinese became vulnerable to the might of arms that finally resulted in the Treaty of Nanking and destroyed the Ching Dynasty and allowed modernity to come in. Thus, they built a wall that prevented the capitalism, science, and technology from coming in. It is a xenophobic fear of new ideas that would undermine the historical traditions of the Chinese people. These attitudes still pervade today in the Chinese stance toward the Internet. They have tried to prevent the invasion of the world by enabling a "Chinese firewall" and "packet sniffers" that prevent and protect the Chinese world from outside intrusion and dissemination of revolutionary ideas into China (Lovell 2006: 296).

In conclusion, the theory of geographic causation has been criticized by many historians, especially Diamond's theory in the *Guns, Germs, and Steel*. Many valid criticisms come from the analysis that the geographic factors such as the east-west axis, the mobility of products, and the domestication of animals have been criticized by several historians. Furthermore, the second group of criticism has focused on the cultural factors, such as Confucian hegemony, social structure, and the Chinese antiforeign feelings, which have bought about serious questions concerning Diamond's thesis. Also, there are several historians who explain that the Europeans exploited the introduction of silver and coal sources, opening markets and trade and increasing the likelihood of capitalism occurring. Finally, my thesis that the building of the Great Wall of China and the interactions with the Mongolians led to an antiforeign attitude as well as the indoctrination of Chinese cultural superiority, leading to the delay in Chinese advancement versus European advancements, was another factor that Diamond did not advance in his discussion. The Great Wall leads to psychological walls that would isolate and still isolate China today, as we can see with communist attitudes toward foreigners. We can only hope that some historian will write a more objective analysis of the progress and advancement of civilizations than Diamond's in the future.

References and Further Reading

Blaut, James M. "Environmentalism and Euro-Centrism." *Geographical Review* 89, 3 (1999).

Blaut, James M. *The Colonizer's Model of the World: Geographical Diffusionism and Eurocentric History*. New York: Guilford, 1993.

Blaut, James M. *Eight Eurocentric Historians*. New York: Guilford Press, 2000.

Diamond, Jared, *Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 Years*. London: Vintage, 2005.

Gernet, Jacques. *A History of Chinese Civilization*. New York: Cambridge University Press, 1982.

Jarosz, Lucy. "A Human Geographer's Response to *Guns, Germs and Steel*: The Case of Agrarian Development and Change in Madagascar." *Antipode* 35, 4 (2003).

Kapstein, Ethan B. *Sharing the Wealth: Workers and the World Economy*. New York: Norton, 1999.

Landes, David S. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*. London: Abacus, 1999.

Lovell, Julia. *The Great Wall: China Against the World, 1,000 BC–AD 2000*. New York: Grove, 2006.

McNeill, William H. *The Rise of the West: A History of the Human Community*. Chicago: University of Chicago Press, 1963.

Needham, Joseph. *Science and Civilization in China: General Conclusions and Reflections*, Vol. 7. New York: Cambridge University Press, 2004.

Pomeranz, Kenneth. *The Great Divergence: China, Europe, and the Making of the Modern World Economy*. Princeton, NJ: Princeton University Press, 2000.

Temple, Robert. *China, Land of Discovery and Invention*. Somerset, UK: Patrick Stephens, 1986.

Walton, Andrew. "A Comparative Assessment of Jared Diamond's Explanation of Inequality." PhD thesis, Politics and International Science, University of Warwick, UK (2006). www.inspire.org/articles/awalton010207_review_essay_diamond.pdf (accessed May 25, 2010).

Wong, R. Bin. *China Transformed: Historical Change and the Limits of European Experience*. Ithaca, NJ: Cornell University Press, 1997.

6

The findings of Neolithic drawings at Çatalhöyük in Turkey are a fraud. (Heated scholarly debate over the authenticity of the Neolithic interpretation.)

PRO	Justin Corfield
CON	Harald Haarmann

PRO

The village of Çatalhöyük (or Çatal Hüyük) in Turkey is close to the city of Konya and south of the Turkish capital of Ankara. There, archaeologists found some 14 layers of mud-brick dwellings, and these are regarded as clear evidence of the oldest known human community, with the earliest remains variously dated to 6000 BCE. Little was known about the early human life in the region until the British archaeologist James Mellaart started excavations there between 1961 and 1965, and the information discovered at Çatalhöyük was dramatic. The archaeological reports were detailed in a large number of journal articles, many in *Anatolian Studies: Journal of the British Institute of Archaeology at Ankara* (an annual journal that recorded in great detail the archaeological work at Çatalhöyük) and a number of books. However, it was not long before Mellaart found himself banned from excavations in Turkey. Indeed he had earlier been expelled from the country due to suspicion of his involvement in the antiquities black market. From an historical angle, Mellaart also became involved in an academic dispute over his ideas concerning a Mother Goddess, and his four-volume *The Goddess from Anatolia*, written in collaboration with Udo Hirsch and Belkis Balpinar, which was published in 1989, probably in Italy, highlights the details of the Neolithic drawings at Çatalhöyük and their link to his new ideas.

In spite of many people working at the site, James Mellaart remains the man most identified with the finds at Çatalhöyük, and he is the author of much of the material about them as well as the source of the majority of the controversy. He was born on November 14, 1925, at the family home in Oxford Street, London, the son of Jacob Herman Jan Mellaart and his wife Apollonia Dingena "Linn" Van Der Beek, both from Hillegom, the Netherlands. Jacob Mellaart was himself a scholar, writing an important booklet *Dutch Drawings of the Seventeenth Century*, which was published by Ernest Benn in London in 1926, and he also collected Chinese antiquities. James's younger sister Helen,

127

The Mother Goddess

Mellaart's findings at Çatalhöyük have been among those used to support various theories about matriarchal or goddess-worshiping Neolithic cultures. The frequent discovery of feminine figurines, often heavy-set, large-breasted, and wide-hipped, has persuaded some researchers that the earliest European cultures revered the feminine more than the masculine. Archaeologist Marija Gimbutas has controversially described an "Old Europe" of matrilineal, peaceful, cooperative, goddess-worshiping communities that were later overtaken by the warring patrilineal tribes of the Indo-European cultures and the Kurgan invasion. There is little firm evidence one way or the other and no way to know whether the figurines are meant to represent a goddess, or even whether their apparent outnumbering of male deity representations really indicates a superiority of one gender over the other.

---

born in 1930, also in London, later moved to the Netherlands where she became an important author.

James Mellaart studied at University College, London, gaining his bachelor of arts with honors in ancient history and Egyptology in 1951. He then went to Turkey, where he was involved in archaeological fieldwork and conducted field surveys with the British Institute of Archaeology at Ankara until 1956. It was his work in 1954 on the "champagne glass" pottery of Western Anatolia during

Aerial view of the archaeological site at Çatalhöyük, a Neolithic farming settlement. Discovered in the 1950s, the ruins are in present-day Turkey on the Konya Plain. (Yann Arthus-Bertrand/Corbis)

the Late Bronze Age that led to the discovery of the site of Beycesultan. However, at the start of the dig, Seton Lloyd, the director of the British Institute, was injured in a car accident when traveling to the site, leaving Mellaart to start the work that helped clarify the nature of life at the time of the Trojan War. However, Mellaart's early conclusions came under attack from Ronald A. Crossland of King's College at the University of Durham, an expert in Hittite philology.

In 1957 Mellaart started work on the site at Hacilar. His first major controversy came in 1958 following his work on a site at Dorak. He reported on some finds, and the details of them appeared in the *Illustrated London News*. The Turkish authorities were concerned as they did not know about these "finds," and Mellaart said he had been able to draw them after seeing them at the house of a young woman named Anna Papastrati, a Greek who lived in the port city of Izmir (formerly Smyrna). The police were unable to find any woman of that name, and Mellaart was immediately suspected of smuggling antiquities and expelled from the country. At the height of the controversy, the Turkish press accused Mellaart of taking the "royal treasure of Dorak" out of the country and selling it for a reported $48 million. This would later lead to the *Sunday Times* reporters Kenneth Pearson and Patricia Connor writing *The Dorak Affair*, in 1967, much like an Eric Ambler spy story, in which they concluded that there was no evidence against Mellaart, although they did unmask much smuggling of antiquities from Turkey, often through a U.S. airbase in the country.

However, Mellaart was able to return and at the end of 1960 he became interested by the newly discovered site at Çatalhöyük, where he worked from 1961 until 1963—while also lecturing at Istanbul University—returning in 1965. In 1964 he was appointed as lecturer in Anatolian (or western Asiatic) archaeology at the Institute of Archaeology at the University of London, a position he held until his retirement in 1991.

The author of many books, James Mellaart's first was *Anatolia before c. 4000 B.C. and c. 2300–1750 B.C.,* published in 1964 as part of Cambridge University's history of the ancient world. He then wrote *Earliest Civilizations of the Near East* (1965), with French and Dutch editions in 1969 and a Portuguese edition in 1971. This was followed by *The Chalcolithic and Early Bronze Ages in the Near East and Anatolia* (1966); *Çatal Höyük: A Neolithic Town in Anatolia* (1967; German edition also in 1967; French edition in 1971); *Excavations at Hacilar* (1970); *The Neolithic of the Near East* (1975); *The Archaeology of Ancient Turkey* (1978); and *Çatal Höyük and Anatolian Kilims*, which was the second volume of the controversial, *The Goddess from Anatolia* (1989).

When work started on Çatalhöyük in 1961, it became clear that the site was very extensive—covering some 32 acres—making it the largest Neolithic site yet uncovered in what was then known as the Near East. The work there quickly took over from his earlier time at Hacilar. Çatalhöyük was quite clearly the most important town in the region at that time, as all the other known

settlements are the size of villages. Volcanic stone dominates much of the geology of the region, and Çatalhöyük is in one of the most fertile positions in the region, almost certainly contributing to its location and size in Neolithic times.

Even by the end of the first archaeological excavations that took place from May 17, until June 29, 1961, it was quite clear the enormity of the Neolithic settlement with which the archaeologists were dealing. Further work was carried out in the autumn, and it was during this time that the drawings were discovered. By the time work was completed in 1965, it was found that there were 14 layers of buildings, with the real possibility of further prehistoric sites beneath them. The wealth of artifacts found meant that it was one of the most important archaeological excavations in the region.

There is no doubt about the importance of the finds at Çatalhöyük. The chipped stone industry was clearly well advanced and one of the best developed in Neolithic Anatolia. Farming implements as well as weapons (for hunting and for war) were found, which clearly shows that this was a flourishing local industry. There were also tools and jewelry made from bone and extensive finds of pottery, including coil pots. Some of the terracotta pots also had geometric designs stamped on them, and it has been suggested that this might show a level of private ownership. There were also many figurines, both of humans and of animals, the latter with some "stab marks," suggesting the possibility of magic hunting rites. However, much attention has been focused on the paintings at Çatalhöyük, and these have been controversial in both whether or not some of them are genuine and, if they are genuine, whether or not they might have been incorrectly "reconstructed." There was much controversy during the mid-1960s, and as a result, the Turkish government closed work on the site at Çatalhöyük and banned Mellaart from returning to the country, although a press report in *The Times* on July 31, 1971, indicated that the ban had been lifted as their Ankara correspondent was able to contact Mellaart by telephone in Istanbul. It was at that time that a Turkish peasant was arrested for forged antiquities. He had worked on the Hacilar dig with Mellaart, but there was never any suggestion that the two had connived on anything. Indeed Mellaart had his suspicions over the peasant; but it reopened a number of controversies involving Mellaart. It was not until 1993 that the site was reopened for archaeological work, which was carried out by Ian Hodder of the University of Cambridge.

The wall paintings around the site of Çatalhöyük are limited to a relatively small number of buildings. These were found after the end of the main excavation period that summer, but this was not entirely suspicious. They had not been visible after the initial "dig." This was because the walls of many of the buildings in Çatalhöyük were covered with a cream-colored plaster. In normal archaeological work, the existing plasterwork would not have been touched, or would have been treated rather than removed. However Çatalhöyük was not like any normal site. The aim for the archaeologists was to dig down below the first

level of buildings they found to find traces of earlier sites there. Thus after one level was excavated and mapped, it was removed for work to begin on the lower level. This, therefore, involved removal of most of the earlier finds. In doing this, some of the plaster was removed from the walls, and, according to the published accounts of the excavation, it was only then that some of the paintings were found. In 1962 these paintings were hailed as being the earliest that had been found on man-made walls, yet their presence on two buildings in Level III and Level IV, and also in two private houses in Level VI, has provoked some intense scholarly debate.

The paint used on them varies. In some places it is just red paint, in other areas the pictures are in a number of different colors. The first major query was why the paintings were only found in such a small number of places. There were also queries as to whether there might have been paintings on other walls that did not survive and whether there was a special reason for these walls to have paintings on them. The latter seemed to point to the idea that a small number of the rooms that were excavated might have been prayer rooms or the location of shrines. Mellaart was quick to conclude that the paintings were not done solely for artistic reasons, but there was clearly some other reason for them. The style of the paintings also varies, and this seems to back his explanation. Some of the paintings appear to have a narrative style approach, with scenes replete with humans and animals, while others have geometric patterns. In the latter, there are images of spirals, chevrons, dots, spheres, and honeycombs, some quatrefoils, and other designs. There are also representations of handprints on them.

Mellaart also revealed that there were a few places where there were successive layers of paintings, separated by white plaster. And it was clear that all the paintings were subsequently covered in white plaster at some stage before the settlement was abandoned. It is easily possible that there were other paintings that did not survive or were not found. Paint samples were also sent to the Courtauld Institute in London to report on the pigments being used. Mellaart was later quoted as stating that the paints were made from minerals or vegetables, and some of the people may have been originally pink in color, but over time the paint had turned brown or black.

The paintings were first uncovered in November 1961 by Mellaart, but the initial problem was that the images had not only been covered by centuries of grime but also with plaster from later use of the rooms. As a result, Mellaart recorded that he had to clean the walls with help of Anne Louise Stockdale, who, Mellaart notes in his 1962 article in *Anatolian Studies*, then copied the paintings. Mellaart and Stockdale received help from Perry Bialor, an American anthropologist, and Mellaart's Turkish-born wife, Arlette Meryem (née Cenani).

Bialor, born in 1931, had completed his graduate work at the Graduate School of Arts and Science at Harvard University in 1957–1958, and then graduated with an AM from the University of Chicago in 1958. In 1973 he completed

his PhD thesis at the University of Chicago, titled "A Century and a Half of Change: Transformations of a Greek Farming Community in the Northwestern Peloponessos, Greece," and moved to live in Astoria, New York. By contrast, Stockdale, born in 1938, was the only daughter of Sir Edmund Villiers Minshull Stockdale, 1st Baronet, former assistant principal at the Bank of England, and his wife, Hon. Louise Fermor-Hesketh, elder daughter of Lord Hesketh. Just after she completed many of the drawings, in 1963, she married Charles Marcus Edwards, a British diplomat who had just finished a posting in South Africa and was to serve in Laos before retiring from the British Foreign Office to practice law and become a judge; she died in 1970.

According to the published material in *Anatolian Studies*, Mellaart's wife and Stockdale were involved in removing the paintings from the wall, preserving them, and then taking them to the Archaeological Museum at Ankara. This was important because, as mentioned before, many of the houses had to be taken down in order to reveal information about the lower levels of the settlement. Much of the site was photographed prior to this, but due to the cost of photography in those days, not everything was photographed, although they were sketched, and extensive drawings were published in various issues of the journal *Anatolian Studies* as well as in *The Times* on February 16, 1966.

Because some of the wall paintings did not survive the excavations and the move, it was not long before skeptics were starting to query their existence, especially the ones for which no photographs seem to have survived. Some claimed that the paintings were entirely faked, but others suggested that although there might have been images and paint marks on the walls, the drawings were sometimes a little too well done, and it was possible that the artists might have interpreted a little more than was actually present, possibly even being too influenced by work that might have been found on other sites. It seemed fairly certain that some of the paintings were genuine (a number have survived and are held in the Archaeological Museum), but what happened to the others?

This led some critics to ask to inspect the actual sketches made at the archaeological site in the early 1960s. Although drawings were published in *Anatolian Studies*, these were clearly done in a studio, and skeptics tried to locate the original sketches, tracings, or indeed photographs that would show what actually been found and what was interpretation or extrapolation from what was found. That reconstruction was carried out was quite clear when comparing the original drawings in *Anatolian Studies* with Mellaart's 1989 book *The Goddess from Anatolia*.

On the face of it, the drawings found are quite spectacular. This is not because of their artistic merit but rather because of what they show. There are only a few color photographs of the drawings and a number of black and white ones. However, most of the illustrations, which were published in *Anatolian Studies* in 1962, have a caption including the phrase "tracing with reconstruction by A. L.

Stockdale" or "after preliminary copies by A. L. Stockdale." Certainly Stockdale is an accomplished artist. That has never been in doubt, and Mellaart noted that the reconstructions were done in the Bosphorus or in England. However, the phrase "after preliminary copies by" is in itself curious. Anne Louise Edwards, as she was then, was still alive, and the phrase was obviously chosen carefully, seemingly to state that the process involved was the location of the painting. Stockdale later traced the image and turned them into a painting or reconstruction either in Turkey or after her return to England, and then a later painting was made of the work by Stockdale, and this was what was being published. The published drawings are clearly much easier for historians to extrapolate information from than the few surviving photographs of the actual paintings themselves. This has led to suggestions that the published drawings show, at best, far too much "reconstruction." The three color photographs in *Anatolian Studies* in 1962 show the outline of what has been surmised to have been a deer hunt (plate 17), but opposite it the reconstruction shows far more detail showing that the published reconstruction had had some elements in the picture surmised. This is quite easy because the two were published opposite each other. Yet for many of the other paintings, color photographs do not exist, and in many other instances, no photographs were published, and probably no photographs were taken. Indeed Mellaart noted that for the earliest wall paintings "it was impossible to detach them from vitrified walls they were covered up after recording" (Mellaart 1962). Stockdale died in 1970 just after the controversy had reached epic proportions in the academic community, but also before the publication, during the 1980s, of the most controversial "reconstructions."

Therefore, a number of archaeologists, historians, and commentators have challenged the nature of these reconstructions. To understand them, it was necessary to discuss the paintings one by one. The paintings from Level VI come from three panels in a private house. All these are geometric designs and are in a number of colors. Unfortunately only copies of drawings survive. None of them are complete, showing that the images found were not complete either. This may have been because they did not survive the rigors of time, or it might have been that some parts were removed before plaster was placed over them in prehistoric times—of course, they were genuine. However, there have been subsequent suggestions that what was found was even more fragmentary than what was shown in the reconstructions. On plate 13, published in *Anatolian Studies* in 1962, from a different building, it is clear that the method used to copy the image included a tracing of it, so there is every likelihood that this is what took place here.

The local people used geometric patterns on pots, so the idea of them doing so on wall paintings is not surprising. To many casual observers, it may seem as though the patterns are not that dissimilar to some modern wallpaper and served to lighten the room. However whether the patterns owe something to symbolism, and their covering up with later plaster was just a change in fashion, can only be surmised.

The real controversy about these paintings from Level VI is whether they show representations of an Anatolian *kilim*, or tapestry. This was suggested by Mellaart but is hotly criticized by his detractors. Marla Mallett, in her article in *Oriental Rug Review* in 1990, claims that this is not the case. In particular she casts doubt on whether the painting shows a particular system of weaving using overlapping forms. Mellaart felt that the patterns were essentially evidence of particular ways of weaving cloth at the time whereas Mallett disagrees. However, her main dispute is not with the reconstructions for which photographs have been published, but on those where it has not.

With no figurative paintings found on Level V, in a Level IV shrine some paintings survived, albeit damaged by the roots of plants and the burrowing of ground squirrels and jerboas. It was on this level, under layers of plaster, that drawings of human figures were found, with women in red and men in white, a convention that was also found in Egypt. Mellaart wrote that women were shown as plump, whereas men were shown as thin. It is possible that this was because the large female figure is associated with Mother Earth, and certainly the image is not that dissimilar to the surviving part of a Neolithic statue at Tarxien in Malta. Again this picture is one in which a photograph of the original wall painting and the reconstruction have been published side-by-side. However, the only images known of the other paintings come solely from reconstructions, and these are so fragmentary that it is impossible to surmise exactly what the scene depicted. Prior to reconstruction, the actual images might have been even harder to ascertain.

This problem about the lack of published photographs of the actual wall paintings is even more important for those in building Level III. Given that many of the reconstructions were done away from the site (a number in England), it is clear that there must have been tracings or photographs unless the images themselves were reconstructions based on only small fragments. The image that is claimed to be of a bearded man is quite clearly open to different interpretations, especially because it is so different from the rest of the paintings' style, which shows narrow figures. Even Mellaart notes that one scene, possibly of a bird or bear trapped in a net, "defies interpretation."

Mellaart noted that the largest number of paintings was from Level III, where a room was found that had paintings on three of its four walls. There are scenes of the hunting of a herd of red deer—some with antlers—that are known to have lived in the region at the time. Another scene shows the hunting of a bull. Again much of this is based on reconstructions, which may or may not accurately record the scenes uncovered.

Mallett's particular criticism is reserved for a number of reconstructions that have appeared many years after the original archaeological work. This was particularly true for some of the images of the "Mother Goddess," and Mellaart's championing of this idea evoked protests from many archaeologists who wanted

to see any surviving photographs of the Neolithic paintings in situ. Mallett wrote, many years later, that she was also unable to locate any of the original tracings and was critical of the reliance, on trust, of reconstructions. For instance, in *Early Turkish Tapestries*, published in 1984, the "bird carriers" and "elibelinde figures" appeared for the first time as reconstructions. Her argument is that these appeared some 20 years after they were said to have been found (and destroyed), but supportive photographs are needed if one is to believe that these did exist.

To illustrate her point, Mallett studied the technology used in the production of textiles at Çatalhöyük versus the conclusions drawn by James Mellaart. Certainly the few surviving fragments of cloth recovered show that an elementary loom was used. However, Mellaart claims that a warp-weighted loom was also used, and this can be seen in the paintings in Level VI. Mallett feels that there is no evidence that warp-weighted looms were used until much later, and that in a 1962 report by Hans Helbaek, cited by Mellaart, there was no reference to this. Furthermore, some geometric patterns were interpreted by Mellaart as clearly showing the weaving of cloth, whereas Mallett believed that the overlapping pictures were geometric patterns, which she suggested could be inspired by "simple basket or rush mat interlacings" (Mallett 1990).

In conclusion, it can be seen that there has long been controversy about the career of James Mellaart as an archaeologist. Some of these controversies have been used to portray him in a bad light, but others have resulted in him being associated with incidents with which he was not really connected. However, the Turkish government did think so poorly of him that he was banned from the country for a period, perhaps unfairly. There is no doubt that Mellaart did locate tremendous finds at Çatalhöyük, that they did depict the early domestication of animals, and that some of the wall paintings do show hunting scenes.

However, the problem has been that a large number of details on many of the wall paintings seem to have only been published in reconstructed form, and it has therefore been impossible to determine whether the resultant paintings are accurate representations, or whether they contain more information and interpretation than were actually found. Marla Mallett's comparison of his early work with the pictures in *The Goddess from Anatolia* certainly highlights discrepancies. And the artist who drew the early reconstructions died in 1970. In *The Goddess from Anatolia*, much more detail is included in updated reconstructions, and this has led to criticism over Mallaart's ideas over a "Mother Goddess" because he has only published one photograph of the original wall painting that has been published and it is not very clear. James Mellaart has long had opportunities to publish some of the tracings that were made of the discovered wall paintings as well as any surviving photographs, but in spite of a long (and distinguished) series of publications to his name, he has declined so to do. For that reason many academicians doubt the authenticity and accuracy of the reconstructions, some going as far as alleging that they are, in fact, fraudulent.

References and Further Reading

Balter, Michael. *The Goddess and the Bull: Çatalhöyük: An Archaeological Journey to the Dawn of Civilization*. New York: Free Press, 2004.

Frauenknecht, Bertram. *Early Turkish Tapestries*. Nürnberg: Verlag Bertram, 1984.

Mallett, Marla. "A Weaver's View of the Çatal Hüyük Controversy." *Oriental Rug Review* 10, 6 (August/September 1990). Also at http://www.marlamallett.com/ch.htm (accessed May 26, 2010).

Mallett, Marla. "An Updated View of the Çatal Hüyük Controversy." *Oriental Rug Review* 13, 2 (December/January 1993a). Also at http://www.rugreview.com/orr/132marl.htm (accessed May 16, 2010).

Mallett, Marla. "The Goddess from Anatolia: An Updated View of the Çatal Hüyük Controversy." *Oriental Rug Review* 13, 2 (December/January 1993b). Also at http://www.marlamallett.com/chupdate.htm (accessed May 26, 2010).

Mellaart, James. *Anatolia before c. 4000 B.C. and c. 2300–1750 B.C.* Cambridge, UK: Cambridge University Press, 1964.

Mellaart, James. "Anatolia before the Indo-Europeans." *Journal of Indo-European Studies* 9, 1–2 (Spring/Summer 1981): 135–49.

Mellaart, James. *The Archaeology of Ancient Turkey*. Totowa, NJ: Rowman and Littlefield, 1978.

Mellaart, James. *Çatal Höyük and Anatolian Kilims*. N.p.: Eskenazi, 1989.

Mellaart, James. *Çatal Höyük: A Neolithic Town in Anatolia*. London: Oxford University Press, 1967. German edition also in 1967; French edition in 1971.

Mellaart, James. *The Chalcolithic and Early Bronze Ages in the Near East and Anatolia*. Beirut: Khayats, 1966.

Mellaart, James. *Earliest Civilizations of the Near East*. London: Thames and Hudson, 1965.

Mellaart, James. "Excavations at Çatal Hüyük," *Anatolian Studies: Journal of the British Institute of Archaeology at Ankara*. 12 (1962): 41–66.

Mellaart, James. *Excavations at Hacilar*. Edinburgh: British Institute of Archaeology at Ankara by Edinburgh University Press, 1970.

Mellaart, James. *The Goddess from Anatolia*. 4 vols. N.p.: Eskenazi, 1989.

Mellaart, James. *The Neolithic of the Near East*. London: Thames and Hudson, 1975.

Mellaart, James. "New Light on Early Farm Communities," *The Times* (July 10, 1963): 8.

Pearson, Kenneth, and Patricia Connor. *The Dorak Affair*. London: Michael Joseph, 1967.

CON

The case of the wall paintings at Çatal Hüyük is special because their authenticity has been debated not only in expert circles of archaeologists and culture historians but also among a broad public, mainly because the sensational finds of the 1960s raised much public interest from the earliest excavations of the site. The controversy over the works of mural art is not a matter of a simple pro and con of whether they are all reality or all a hoax. Some paintings undoubtedly are authentic because the originals have been preserved, photographed, and documented, and others are controversial because no originals are available and the nature of their narratives differs markedly from those of the authentic category.

The images of the authentic category will always retain their value, while the others may always remain controversial. The crucial issue of this controversy is how to make a distinction between the two categories of paintings. The Romans had a saying: *Nomen est omen* ("fate or fortune is in a name"). Çatal Hüyük literally means "forked mound" (there are Çatal Hüyük East and Çatal Hüyük West). The dichotomy of authentic and controversial wall paintings may be seen as a reflection of a metaphorical bifurcation of the archaeological recording of antiquities at this site.

The debate about the wall paintings is frustrating because the outsider finds him- or herself in a jungle of contradicting opinions. The history of the discoveries at Çatal Hüyük is an adventure story that is saturated with many mysteries in the biography of the protagonist, British archaeologist James Mellaart, with whose name this Neolithic town is intrinsically associated. His colorful personality and his mysterious encounters have made him a controversial figure. Many appreciate him as an impeccable explorer of humankind's prehistory, while others doubt his credibility as a scientist and even accuse him of fraud.

To find valid answers, the pieces of a puzzle must be put together, and these pieces are quite different in nature. The controversy over the wall paintings has to do with facts and evidence, but also with the image and the reputation of the persons involved in the discoveries. This discussion is an exercise in the psychological undercurrents of the world of excavators and their work.

Çatal Hüyük Site and Mellaart's Discovery

Çatal Hüyük, situated in the Konya Plain of southern Anatolia (Turkey), was discovered in 1958 by a group of archaeologists led by Mellaart. The most recent calibrated dating indicates that the site was inhabited between ca. 7400 and 6000 BCE. During four seasons of excavation, between 1961 and 1965 (with an interruption in 1964), Mellaart's team uncovered over 300 rooms, many of which contained not only well-built domestic features and skillfully produced

crafts and tools, but also dozens of polychrome wall paintings, clay and plaster bas reliefs, ritual installations, and sculptures, indicating a pervasive and sophisticated ritual life that extended throughout the duration of the Neolithic occupation. The people of Çatal Hüyük were described by Mellaart in *Çatal Hüyük and Anatolian Kilims* as being skilled in agriculture, architecture, painting, weaving, and mortuary arts (Mellaart 1963).

The discovery of Çatal Hüyük was undoubtedly the highlight of Mellaart's career. With his narrative skills, he succeeded in spreading the sensational news about the oldest large agricultural settlement of the world, which had the size of a small town and an estimated population of 6,000 to 8,000 inhabitants. Mellaart and his team unearthed treasures of world cultural heritage. As the leading excavator, he earned a worldwide reputation as a great discoverer. And yet at the height of his professional career, Mellaart stumbled over his own feet and collided with the Turkish authorities who had granted him the permission to dig at Çatal Hüyük.

What made Mellaart stumble and lose his authority as head of the excavation team was not one incident—the suspicion of looting artifacts and smuggling them from the site. It was a series of controversial events in which Mellaart was rumored to have been involved. The image of Mellaart's controversial personality was formed several years before the breakthrough discovery of Çatal Hüyük. At the beginning of the chain of events that finally discredited Mellaart in the eyes of the Turkish authorities was the so-called Dorak affair. When it all started has never been clarified, and Mellaart himself reported various versions of a story that, to many, seems too unusual to be true.

The Dorak Treasure

It was as if Mellaart, when making his career in archaeology, had carried Pandora's box, which he accidentally opened one day when he was traveling by train to Izmir on the Aegean coast. In the train, the British archaeologist met a young attractive woman, Anna Papastrati, who promised to show Mellaart pieces of a lost treasure. Pandora's box released its poisonous contents some time later: doubt and suspicion, scorn and scandal, humiliation and delayed exoneration.

Nobody except Mellaart himself knows the true story, and he has not produced any consistent version of what happened and when it happened. There are certain elements in this mysterious encounter that are uniform in all the narratives and reports about it. Mellaart followed Anna's invitation to her home in Izmir (located in Karsiyaka, a suburb). There he spent an entire week with the young woman. Anna showed him artifacts, and Mellaart made drawings of them. In his accounts, Mellaart gave different dates for his stay in Izmir. Sometimes he said this was in 1958; on another occasion, he mentioned that it took place years earlier, in 1952.

Neolithic Revolution

Çatal Hüyük dates from the first agricultural "revolution," when hunter-gatherer groups began to settle down in large numbers and form sedentary societies with significantly greater population density, based around agriculture and permanent dwellings. Doing so by turns required and enabled a number of significant developments in human prehistory: pottery (including the Çatal Hüyük figurines), barter and trade (such as between a shell-collecting community and a community with access to obsidian), and communal defenses such as village walls. Social developments alone included marriage, property, and the inheritance by children of their parents' belongings, all things that made more sense than they had in a nomadic lifestyle.

When asked why he would have kept such an important encounter and discovery of artifacts a secret for many years, Mellaart gave different reasons. One was holding back information as a protection of the treasure. Another reason given was more personal: Mellaart did not want to stir up any resentment on the part of his wife, Arlette Cenani, whom he had married in 1954.

The exact date of the encounter in Izmir has no bearing on the event that triggered the public scandal. Mellaart published his collected drawings and a detailed, four-page description of what he called a Bronze Age treasure in the *Illustrated London News* in 1959. This publication of extraordinary finds that were hitherto unknown certainly raised great attention in expert circles and among Turkish authorities, who were eager to inspect the original artifacts.

According to Mellaart's reports, Papastrati did not disclose the source from which her family had obtained the Dorak treasure, but she had promised to send photographs and her written permission to publish the drawings. Mellaart claimed that he had received a letter from Papastrati granting him permission to publish pictures of the finds. However, the envisaged photographs never arrived. Something else weighed even heavier and cast doubts on Mellaart's credibility: the original artifacts never showed up and could not be located by the Turkish authorities. This raised suspicions that the Dorak treasure might have been smuggled out of the country.

Mellaart had his reputation at stake. With his successful excavation of the site of Hacilar, which started in 1957, Mellaart had made a name for himself, not only among archaeologists but also among a broader public in Turkey. It was the scandal-hungry Turkish media that stirred up positions pro and con. Mellaart's image of an impeccable scientist suffered because there were many people, especially in Turkey, who suspected the British archaeologist was involved in the alleged "disappearance" of the Dorak treasure. It also became known that some of the artifacts that were auctioned at Sotheby's in London in the late 1950s belonged to the finds retrieved at Hacilar. Here, again, Mellaart's reputation as the leading excavator was at stake.

Although no evidence for Mellaart's possible involvement in the "disappearance" of the Dorak treasure (whether or not it ever existed) and in the smuggling of Hacilar artifacts to Britain was ever presented, the mere uncertainty and suspicion left a shadow that continued to grow. The file of the Dorak affair was finally closed—at least in the eyes of a broader public—when two investigative reporters, Kenneth Pearson and Patricia Connor, affiliated with the *Sunday Times*, published their detailed two-part report in November 1966. They concluded that Mellaart was not intentionally involved in any illegal action relating to Turkish antiquities. The same conclusion was repeated in a book by Pearson and Connor in 1967 in their book *The Dorak Affair*.

These publications on the Dorak affair meant an exoneration of Mellaart's reputation. But the earlier suspicions had already taken their toll. Mellaart had been barred from archaeological excavations in Turkey, and this decision of the Turkish authorities was not reversed. Although the two reporters excluded any possibility of Mellaart's active involvement, they nevertheless brought up another idea that was not favorable either: Had Mellaart perhaps become an unwilling victim of smugglers who wanted to have an expert's opinion on their looted finds? Had Papastrati, the mysterious woman in the train, been an accomplice of the smugglers who was sent to capture Mellaart's attention? Had Mellaart, unaware of Anna's involvement with smugglers and in his scientific perfectionism, elaborated a testimony for the smugglers in his newspaper article?

As things are, there will never be any definite answers to those questions.

Sensational Finds Overshadowed by Scandal

Mellaart was a visionary, and this capacity granted him success with his diggings at Çatal Hüyük right from the beginning. The excavation team that explored the site under Mellaart's direction had great expectations. These had been sparked by the first sensational find of clay figurines at Hacilar in the late summer of 1960. In early 1961, Mellaart had published a report and the photos that his wife, Arlette, had taken of the artifacts in *Anatolian Studies*. Mellaart's interpretation of the figurines as representations of the "Mother Goddess" were readily accepted by the British press and became seminal. In his report, Mellaart expressed his hopes that an even older civilization than that of Hacilar might be found one day that would harbor the early stage of the figurines as a genre of Neolithic art. Mellaart was to become the prophet of his own words.

What he would soon find at Çatal Hüyük was indeed older. The figurines from Hacilar were a sensation. But nobody had any idea that the discoveries of antiquities at Çatal Hüyük that were immanent would surpass the finds of Hacilar in their sensational value. Excavation work at Çatal Hüyük started in May 1961, and it did not take long before an invaluable treasure of world cultural heritage was discovered: wall paintings. The personal credit for the successful

discovery of the first finds of painted images on the walls of the houses at Çatal Hüyük goes undisputedly to Mellaart. The tell of Çatal Hüyük covers a large area (some 32 acres), and only a man guided by his intuition would know where to best start the archaeological explorations. Mellaart had decided to dig into the tell in its southwestern section.

It was on the second day of excavation work that Mellaart's sensitive eye was caught by something unusual in the rubble and the innumerable pieces of plaster that were removed by the spades of the workmen. What might have been overlooked by the untuned eyes of most observers as a formless blotch of red paint on a layer of plaster assumed the contours of an animal in Mellaart's perception: a stag. He started to remove pieces of a layer of cream-colored plaster to uncover what was underneath. The more Mellaart cleared the hidden surface with a small knife the more pictures emerged. After hours of painstaking clearing, a whole scenery appeared, with pictures of hunters chasing deer.

When Mellaart made this discovery that was to become a world sensation, he was not alone. Some 40 pairs of eyes followed his action to separate the layers of plaster at the digging site. There could be no better testimony for the authenticity of what came to light than the presence of the entire excavation team. Arlette's photographs of the images provided solid documentation of these earliest wall paintings in cultural history.

Between 1961 and 1965, several dozen wall paintings were found and photographed, and drawings of them were made. During the first discoveries in 1961, Mellaart noticed that sunlight had a damaging effect on the quality of the colors, on the consistency of the plaster that soon started to crack, and on the overall stability of the unearthed surfaces with paintings. Although conservation work was started immediately and sophisticated measures were taken to have the finds transported to the archaeological museum in Ankara (Museum of Anatolian Civilizations), many paintings disintegrated and only some could be put on display as exhibits. Thus, despite genuine efforts to conserve the precious finds, only a small number of the original paintings could be saved. Conversely, documentation even of the paintings that were lost was secured by the scrutinizing photographs and the drawings of motifs and scenes.

It seems as if Mellaart not only enjoyed the success of the discoveries of wall paintings that his team continually uncovered, but he also somehow became obsessed to unearth ever more paintings. Mellaart was driven by his enthusiasm to find more substantive evidence to prove his hypothesis of goddess worship at Çatal Hüyük. He also had a growing fear of losing permission to continue his excavations. This uncertainty about the attitude of the Turkish authorities caused immense psychological pressure that was to become so typical of Mellaart's last years of digging in Turkey. He suffered growing suspicion of smuggling antiquities (like the imaginary Dorak treasure) or of looting them from their original sites (like some of the figurines from Hacilar).

Mellaart's position as the chief excavator of Çatal Hüyük had been questioned from the beginning. Regardless of his expertise, his intuition, and his good fortune that led him to make sensational finds, his actions always stood in the twilight of old and new suspicions that continually eroded Mellaart's credibility as a charismatic scientist. In the public opinion, suspicions of a personal involvement in the looting of artifacts mingled with blame of a lack of authority on Mellaart's side in his responsibility as the head of the excavation team. From a strictly formal standpoint, Mellaart could be held responsible for any wrong action done by anyone under his direction. And it was a breach of such a formal responsibility that finally brought Mellaart to his knees.

Rumor has it that in the course of 1964—the year Mellaart was refused permission to dig—he had become a liability to the BIAA (British Institute of Archaeology at Ankara) and to the Turkish authorities, mainly because of the bad publicity in connection with the press coverage of the Dorak affair. A lawsuit filed against Mellaart in 1965 was dismissed for formal reasons because—according to an amnesty declared in 1960—he could not be prosecuted for possible illegal actions before that date. Nevertheless, this made Mellaart's position as director of the excavations untenable. He was allowed to continue his work for the campaign of 1965 only as a team member, under the formal direction of Oliver Gurney (the editor of *Anatolian Studies*). This situation was somehow odd, because everybody knew that, in practice, it was Mellaart who was conducting the digging.

In 1961 the excavations had begun with the sensational find of a wall painting, and in 1965 they ended with another such sensation. Mellaart's team retrieved a mural with the biggest picture that has ever been found at Çatal Hüyük, the image of a wild bull in red. The picture of this animal measures some 2 meters (6 feet) in length, within a scene of ceremonial dancing covering a space that extends over almost 6 meters. This scene with the red bull in focus is the most widely known of all the wall paintings at Çatal Hüyük.

Regardless of the fact that the season of 1965 was perhaps the most successful of Mellaart's archaeological campaigns, the joy about the finds was ultimately dimmed by yet another scandal that was disclosed toward the end of the digging season. Instead of the regular representative of the Turkish authorities who was always present at Çatal Hüyük at the time of digging, during the excavations of 1965, two representatives were present for the whole duration of the campaign. In September 1965 one of them detected artifacts from Çatal Hüyük in an antiques dealer's shop in the nearby town of Konya.

The dealer identified the workmen who had smuggled the artifacts out of Çatal Hüyük who then left the site. Although Mellaart was no longer formally in charge, the blame for the smugglers' action was laid on him. This latest scandal meant the end of Mellaart's digging activities in Turkey. He was barred from Çatal Hüyük and was allowed to return only many years later, and only
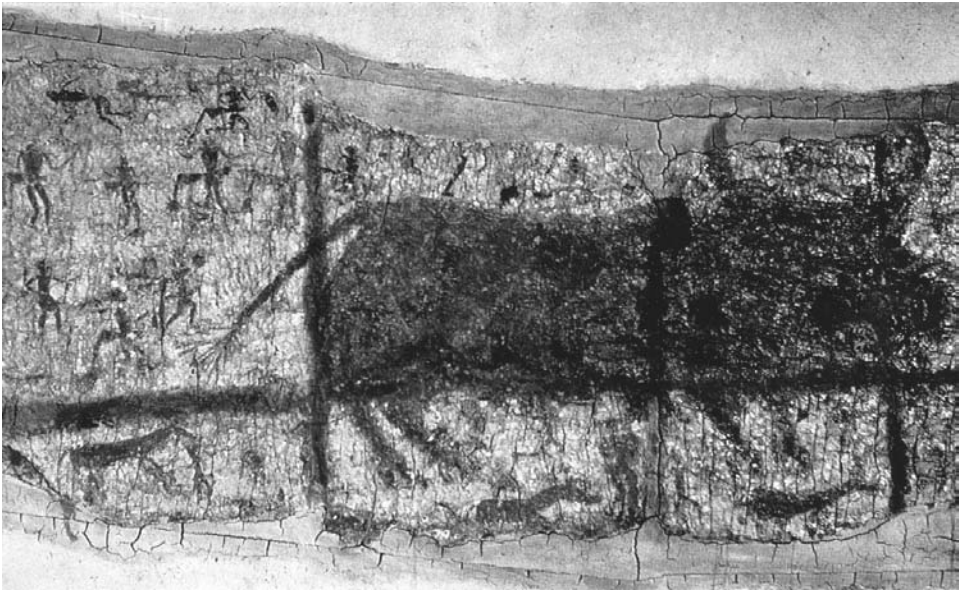
Wall painting from Çatal Hüyük, sixth millenium BCE. (Archiv Gerstenberg/The Image Works)

then as a visitor, not a responsible excavator. The site was closed for decades and reopened for excavations as late as 1993, this time under the direction of the British archaeologist Ian Hodder.

Hodder opposes many of Mellaart's interpretations of the symbolism of female figurines, reliefs, and of communal life at Çatal Hüyük. Hodder, nevertheless, acknowledges the authenticity of the finds that Mellaart made, which are documented in the reports published in the 1960s. Among the pieces of world cultural heritage that continue to be studied as objects of indisputable value are the wall paintings available through photographs and drawings. Arlette took photographs, whereas the drawings were made by Mellaart himself and by some professional artists.

Authentic Record of the 1960s, 1990s, and 2000s

Before the first digging campaign of 1961 had ended, the British press—above all the *Daily Telegraph*—lavishly reported about Mellaart's finds, and the media continued to follow the discoveries that were made by Mellaart and his team during the following years. Among the highlights of the newspaper coverage is the two-part essay, illustrated with color photos, that was published by Mellaart in the *Illustrated London News* in January and February 1963.

As for the scientific discipline of archaeology, the documentary material was presented by Mellaart in articles that appeared in *Archaeology* and *Anatolian*

Studies. These became the preferred forums for Mellaart's early reports on Çatal Hüyük. The sum of his experiences from the four successive excavations between 1961 and 1965 was presented in the first book publication about Çatal Hüyük, *Çatal Hüyük: A Neolithic Town in Anatolia* (1967). Although Mellaart had taken to calling the Neolithic settlement a "city" because of its sheer size already during the first season and in his article in *Scientific American*, he nevertheless decided to use the term "town" in his book.

Mortimer Wheeler, the gray eminence of British archaeology, praised the achievements of the Neolithic settlers in his introduction to Mellaart's *Çatal Hüyük* and assigned Çatal Hüyük a key position in archaeological studies: "[Çatal Hüyük] represents an outstanding human accomplishment in the upward grade of social development, and may be expected therefore to be of general interest even to a modern age which may have lost something of the easy Victorian certainty of Progress" (in Mellaart 1967:33). What Wheeler referred to as "Victorian" was the traditional view of civilization (high culture) having originated in ancient Mesopotamia. Until the discovery of Hacilar and Çatal Hüyük, Anatolia had been considered a backwater of prehistory.

Mellaart's 1967 monograph on Çatal Hüyük became seminal as a basis for numerous scientific studies on the Neolithic community and its material and spiritual culture. There is a consensus among archaeologists and culture historians that Mellaart's book is a reliable source for the study of the material culture at Çatal Hüyük, and this consensus also extends to include the motifs and scenes of the wall paintings that are documented in this publication.

A further acknowledgment of the original drawings of wall paintings from the excavations between 1961 and 1965 is their inclusion in the exhibition "From Earth to Eternity: Çatal Hüyük," held between May 26 and August 20, 2006, at Yapi Kredi Vedat Nedim Tör Museum in Istanbul. Mellaart's sketches and restitution drawings are appreciated as "historical archives and a collection on its own."

The finds of wall paintings did not remain limited to the excavations that were conducted by James Mellaart in the 1960s. New discoveries of mural art were made after the archaeological site of Çatal Hüyük had been reopened in 1993 and excavations had continued under the direction of Ian Hodder. Although not comparable in the magnitude of their sensational value with Mellaart's finds, the fragmented wall paintings that have been retrieved since 1993 nevertheless add to the rich repertory of the visual arts at Çatal Hüyük. In addition to their documentation in excavation reports, collections of the newly discovered mural paintings can be found in such recent books as David Lewis-Williams and David Pearce's *Inside the Neolithic Mind* (2005) and Ian Hodder's *The Leopard's Tale* (2006).

The wall paintings from Çatal Hüyük are polychrome, and individual motifs are depicted in a variety of colors, including white, black, red, pink, orange,

brown, and yellow. The individual motifs found in the paintings belong to the following categories:

- wild animals (i.e., leopard, deer, bull, boar, vulture, crane),
- human beings (i.e., hunters with bows, dancing male figures dressed in animal skin),
- geometric forms (i.e., triangle, square, v-sign, cross, dot, line, grid),
- wild plants (of rare occurrence, e.g., a tree).

As Hodder points out in *The Leopard's Tale*, wild animals, including bulls, deer, and goats, is the main theme in the painting (2006).

Most motifs appear in scenic assemblages as to form narrative compositions. These may be categorized as follows:

- scenes of hunting (involving human figures of hunters and animals),
- scenes of ceremonial dancing (probably in connection with hunting rituals),
- scenes with vultures and headless human figures,
- groupings of animals (e.g., two leopards or cranes facing each other),
- schematized landscape with the layout of the town and a volcano in the distance, and
- compositions of purely geometric motifs.

In the largest image, the red bull in the center of a scene is surrounded by figures that appear to be dancing, some holding bows. One human figure is depicted as kneeling on the back of the bull, while a headless female figure stands in a central position under the great bull. Other allusions to dancing are found in other scenes, like the images of two cranes (in black color) facing each other in graceful postures. Cranes are known for their elaborate mating dances.

Among the exceptional paintings is an assemblage of rectangular shapes with subdivisions marking the interior spaces that give the impression of the layout of the ancient city. In another scene, two leopards face each other. The design on the animals' skin is painted in black rectangular forms filled with a white cross-like motif.

Evidently related to the mythical world of the inhabitants of Çatal Hüyük and, presumably, to their burial customs are scenes with stylized vultures in close association with headless human figures. These settings might reflect the custom of open-air burials, with the defleshing of the corpses by vultures.

In addition to the abundance of naturalistic images, there is a variety of abstract motifs, some with geometrical shapes (i.e., triangles, rhombic forms, zig-zag lines) recurring as ornamental friezes. These motifs may not have functioned

simply as elements of decorative design but may have carried symbolic meanings. This idea is supported by the presence of hand stencils in association with abstract motifs.

The abundance of wall paintings suggests that the narratives had special significance for the inhabitants of Çatal Hüyük. Together with other genres of visual art (i.e., bas reliefs, figurines, figurative images on seals), the wall paintings testify to the vivid cultural symbolism in the community life of the early agrarian settlers and to their spirited worldview.

In the scientific community, there is consensus about the distinctiveness of the wall paintings at Çatal Hüyük and their fabric: there is nothing comparable in Anatolia or the Near East during the seventh millennium BCE; the number of wall paintings found in so many buildings within the same community is unique among all Neolithic settlements; the authenticity of the finds of wall paintings documented for 1961 to 1965 and since the reopening of the site in 1993 remains undisputed.

Controversy over Drawings Published in the 1980s

The excitement about Çatal Hüyük as a Neolithic settlement would have remained untainted with speculations about the authenticity of its distinctive visual art had not an entirely new, hitherto unpublished series of drawings of allegedly original wall paintings appeared in the late 1980s. In a richly illustrated four-volume publication, *The Goddess from Anatolia*, visual motifs from Çatal Hüyük are compared with motifs in Anatolian kilim carpets. Many of the drawings included in that volume had been unknown to circles of experts and to the public.

In these illustrations, referred to here as the controversial series (CS), new motifs appear in a divergent narrative style that is absent from the authentic series (AS)—that is, from the set of indisputably authentic wall paintings documented in the 1960s. For example what strikes the eye of the observer in the images of the AS is the dominance of wild animals. In the CS, domesticated animals (e.g., cattle, sheep, goat) abound. This is strange, because the authentic scenes with wild animals come from the same stratigraphic layers of the settlement as the controversial scenes.

The images and scenes of the CS differ from those of the AS in many respects. The most striking differences are the following:

Thematic focus
AS: hunting scenes; ceremonial dancing; ritualistic juxtaposition (e.g., animals facing each other in pairs).
CS: depiction of domestic activities such as plant cultivation or transport of goods; appearance of the goddess as agent; detailed depiction of cultic assemblages.

Dimensions

AS: the motifs appear seemingly dimensionless in the scenes (e.g., hunters are positioned around animals in a scene as if viewed from the air although depicted from the side); ground and vegetation are not shown; images focus on essential features with many details missing.

CS: all motifs are rigidly bound to three dimensions with an accurate depiction of details.

Mode of visualization

AS: symbolic meanings are expressed by means of allusions and the overall symbolism is implicit in the visual motifs; scenes remain enigmatic.

CS: a marked trend toward explicitness in the arrangement of motifs; scenes have explicitly expressed meaning (e.g., the goddess supervises the offering of a bull; inhabitants of the town are engaged in the sowing of seeds on a field).

The drawings of the controversial series published by Mellaart in 1989 and 1999 stirred up extensive criticism. The criticism was directed toward two targets. First, Mellaart's interpretations of goddess worship at Çatal Hüyük that were (and are) disputed by many. Second, the lack of original documentation of the pictures that Mellaart uses to support his interpretations. The criticism coagulated in some longer articles, the most substantive published in 1990 in the journals *Oriental Rug Review* and *Hali*.

A particular facet of the criticism is the articulated skepticism as to the authenticity of the wall paintings whose compositions the drawings supposedly render. No documentation of original paintings has been produced for any of the drawings of the controversial series. There are those who openly accuse Mellaart of fraud. Others, such as Marija Gimbutas (1990) in "Wall Paintings of Çatal Hüyük," accept the paintings as authentic or think it is improbable that the drawings are a hoax because Mellaart was not aware of the meaning of the cultic symbolism surrounding the figure of the goddess and her attributes.

Mellaart reacted to the criticism in several media, among them *Hali*. He stresses the fact that the originals of many wall paintings that were allegedly documented in drawings that disintegrated soon after having been unearthed, or their pieces were scattered, which made it impossible to photograph them as whole pictures. The question arises: If the state of many wall paintings was that desolate, how could such accurate and detailed scenes be drawn in the sketched reconstructions as they appeared in 1989? Mellaart refers to the fire, in 1976, of the country house (of his father-in-law) where much of the documentary material (photographs and drawings) had been stored. So, the veil of mystery in which the controversial series is shrouded may never be lifted, and suspicions about their authenticity will remain.

Of the drawings of the controversial series, several were republished on different occasions in the 1990s. Among those sources is a longer article by Mellaart that appeared in the Turkish magazine *Cornucopia* in 1999. Here the drawings were presented—without any critical reservation—as visual documents of everyday life in Çatal Hüyük. The controversial drawings have not been accepted for the canon of scientific documentation of architecture, visual art, and community life of the Neolithic settlement at Çatal Hüyük.

References and Further Reading

Balter, Michael. *The Goddess and the Bull. Çatalhöyük: An Archaeological Journey to the Dawn of Civilization*. New York: Free Press, 2005.

Gimbutas, Marija. "Wall Paintings of Çatal Hüyük." Review of Mellaart, Hirsch, and Balpinar. *Review of Archaeology* 11, 2 (1990): 1–5.

Haydaroglu, Mine, ed. *From Earth to Eternity: Çatalhöyük* (exhibition catalogue). Istanbul: Yapi Kredi Cultural Activities, Arts and Publishing, 2006.

Hodder, Ian. *The Leopard's Tale. Revealing the Mysteries of Çatalhöyük*. London: Thames and Hudson, 2006.

Lewis-Williams, David, and David Pearce. *Inside the Neolithic Mind. Consciousness, Cosmos and the Realm of the Gods*. London: Thames and Hudson, 2005.

Marler, Joan, and Harald Haarmann. "The Goddess and the Bear: Hybrid Imagery and Symbolism at Çatalhöyük." *Journal of Archaeomythology* 3, 1 (Spring 2007): 29–60.

Mellaart, James. "The Beginnings of Mural Painting" *Archaeology* 15 (1962a): 1.

Mellaart, James. *Çatal Hüyük: A Neolithic Town in Anatolia*. London: Thames and Hudson, 1967.

Mellaart, James. "Çatal Hüyük in Anatolia: Excavations which Revolutionised the History of the Earliest Civilisations," part I. *Illustrated London News* (January 26, 1963): 118–21; part II (February 2, 1963): 160–64.

Mellaart, James. "Excavations at Çatal Hüyük." *Anatolian Studies* 12 (1962b): 41–66.

Mellaart, James. *The Goddess from Anatolia*, Vol. 2: *Çatal Hüyük and Anatolian Kilims*, edited by Udo Hirsch. Milan: Eskenazi, 1989.

Mellaart, James. "James Mellaart Answers His Critics." *Hali* 55 (February 1991): 54–55.

Mellaart, James. "A Neolithic City in Turkey." *Scientific American* (April 1964): 94–104.

Mellaart, James. "Under the Volcano." *Cornucopia* 19, 4 (1999): 76–99.

Pearson, Kenneth, and Patricia Connor. *The Dorak Affair*. London: Michael Joseph, 1967.

7

The existence of Atlantis is not entirely mythical.

PRO   László Kocsis
CON   Cheryl Golden

PRO

According to many, Atlantis existed in the form of a lost civilization or a vast island empire, which was destroyed suddenly by the end of the last Ice Age, apparently due to geologic factors, namely earthquake and a subsequent tidal wave or tsunami. The cause of these catastrophic events is unclear. In Plato's accounts, there are no mentions of any volcanic eruption that might have been lost in the chain of oral passage of knowledge, but could invariably be the main cause of the cataclysm. The ancient Greek philosopher Plato (427–347 BCE) gives accounts about the empire and its sudden disappearance in his dialogue books *Timeos* and *Critias* (or Atlantic speech) written ca. 360–350 BCE. There he mentions that lying in front of Hercules's columns or "beyond the Pillars of Hercules" (most probably meaning the Straits of Gibraltar, also acknowledged as such by Herodotus) there was a big island with many small islands around, which could have been considered a continent, which may have been called Atlantis. In this view, Atlantis was a confederation of kingdoms and was a naval power that conquered many parts of western Europe up to Tyrrhenia and also parts of Africa 9,000 years before the time of Solon, or by ca. 9500 BCE.

It is worth noting that Plato, in turn, had relied on Egyptian historical accounts that were many thousands of years old, preserved mainly through oral tradition. Plato got the information about Atlantis from Critias, who had obtained those from his grandfather, who had obtained them from the Athenian lawmaker Solon (639–559 BCE), who had been in Egypt in the sixth century BCE. There, the priests of Saida (Sais), the ancient religious capital of Egypt, revealed to him the ancient secrets of Atlantis, written in hieroglyphic symbols on papyrus, by translating them into Greek as accounted for by the ancient historian Plutarch in his account of Solon's life. The most learned of all Egyptian priests had told Solon that the Aegean islands, Crete, and almost all the coastal areas of the Mediterranean countries from north to south reaching up to the Persian Gulf had been conquered by the Atlantes, coming from the big ocean in the west. Proclus, an indirect second-generation student of Plato, reported that Crantor traveled to Egypt and actually found columns with the history of Atlantis written in hieroglyphs at the temple of

149

Niith, which told him that "during the very old times there was a huge island in the Atlantic Ocean" that was much bigger than Libya and Asia together. It must be mentioned that in antiquity the term Libya encompassed all of North Africa outside Egypt, and stretched west as far as to Mauritania, while Asia, in the Greek view, had meant the Asian empires of Persia, India, and China. It is worth noting that the Phoenicians had reportedly sailed west that far, and possibly also reached South Africa and the Bahamas. On the other hand, the Vikings probably reached North America through their voyage via Iceland and Greenland, so the physical connections between the Americas and the Old World, like the Mediterranean, cannot be ruled out prior to the discovery of Columbus, rather than the knowledge of these voyages being lost. It must be mentioned that the large time periods between the creation of these different accounts makes them susceptible to alteration and misinterpretation, and major translation errors cannot be ruled out as well.

It should be mentioned that the claims of the existence of Atlantis may also refer to other civilizations or lost continents that had disappeared, such as Mu and Lemuria, and there are many who consider it also as embodying the upper-level knowledge that had somehow vanished before the modern civilization started to emerge. Some even consider the Atlanteans extraterrestrial in origin, and others as the ancestors of the master or root race or "Nordic-Atlantean" or "Aryan-Nordic" supermen.

But Plato's account may also be considered a political manifesto, which claims the superiority of the Athenian democracy that won over the tyranny represented by the antagonistic society of the Atlantidian Empire. According to Plato, the Atlantes were unable to conquer Attiki because their troops had been defeated by the Athenians. Then, according to the story, Atlantis disappeared suddenly but, interestingly, after a failed attempt to invade the allegorical perfect society of ancient Athens, so it is considered by the historian as the punishment of the gods. Atlantis sank into the ocean in a single day and night of misfortune. Although the fact that the function of the story was to convey Plato's political theories seems clear to many scholars, there is vivid research and much dispute over the veracity of the accounts and their possible basis and the possibility of inclusion of memories of major past political or geological events. A great number of scholars were or are admitting the past existence of Atlantis, and many are actively searching for its possible locations. Many theories have been produced and in order to support the claims for its existence, they use etymological, mythological, geological, and archaeological evidences.

Etymology of Atlantis and Linguistic Considerations

Atlantis is Greek for the "islands of Atlas," and their location seems in concordance with other facts: south of the strait lies the mountains of Atlas or anti-Atlas

in northern Africa, mainly in Morocco and Algeria, which were named for the Greek god Atlas. However this is not the only possible etymology of Atlantis. "Atl" means water in Aztec dialects, Atlaua signifies the water god, which is the same as Atlas in Greek mythology. Atlatonan (or Atlatonin) represents the goddess of the coast, which is the closest match to Atlantis and the Atlantic Ocean, showing clearly the logic of denomination. Derived from Atla comes also Atlacamani, the goddess of oceanic storms, and Atlacoya, the goddess of drought, showing that Atla(s) is not a loan word of the Nahuatl language. Tlaloc, the Aztec god of rain and fertility, may also be a derivative of Atla(ua). Other gods' names like Quetzalcoatl, the feathered flying serpent or Mixcoatl also contain the "atl" root. Historical accounts also mention the tribes of the Aztlans as being the ancestors of Aztecs and Toltecs. Donnelly (1882) claims that the Aztecs pointed east to the Caribbean as the former location of Aztlan, and this may also assert a connection between their ancestral home and Atlantis.

Other Mesoamerican gods' names also contain -atl or even more, like the Mayan word Atitlan from Guatemala, which represents similar mythology. In the local Mayan dialect the word Atitlan translates as "the place where the rainbow gets its colors." Atitlan, in reality, represents a 600-meter-deep, beautiful volcanic lake, and also a volcano, both geologic formations suitable for the myth. The lake is situated 1,500 meters above sea level, surrounded by three majestic volcanoes, one of which is the 3,535-meter (11,598 feet) high Atitlan. Etla, Tlakopan, Tenochitlan, and Mitla are all derivatives of Atla, itla, etla, depending on the deciphering of the writing and its transcription in the languages of the historical accounts. Mixcoatl and Tlaloc and, interestingly, their traditional giants, Atlan and Teitani (Titan), are known from the very ancient times around the Mediterranean as well.

Many linguists claim that Europe was inhabited by pre-Basque peoples prior to the arrival of the Indo-Europeans, who squeezed out these agglutinative languages, their survivors being the Basques, Hungarians, Finns, Estonians, and Turks in Europe, but most of the extinct languages were of this type like the proto- or Volga-Bulgarian, Hun, Petcheneg, Cuman, and Etruscan, among others. These languages, along with the other Ural-Altaic languages like the Kazakh, Turkmen, Azeri, Uzbek, Uyghur, Mongolian, Korean, and Japanese, are all agglutinative type of languages like the Quechua and Aymara (the main languages spoken by the descendants of the Incas in South America) having a similar type of logic and word formation.

It is a fact that many cultures around the world, from India, Sumer, Egypt, Peru, the Indians of North and Central America, the Inca and the Maya, and the Szekler Hungarians or Kazakhs, call themselves the "children of the sun" or the "children of the light." Anthropological and genetic evidence suggests that many speakers of these languages are related to one another; for instance, the descendants of the Maya, the isolated community of the Lacandon Indians in Chiapas, Mexico, show a surprising similarity to the Basque and Berber peoples.

Mythological Parallels between the Old World and the New World (Americas)

Atlas was the name of the Greek god, a primordial Titan who supported the heavens in the form of a celestial globe, rather than the Earth, as it is considered incorrectly. Atlas was the son of the Titan Iapetus and the Oceanid Asia, one of the daughters of the ocean. In Greek and Roman mythology, the Oceanids were the 3,000 daughters (nymphs), or other accounts also include the 3,000 sons, of the Titans Oceanus and Tethys, each of them being the patroness of a particular spring, river, ocean, lake, pond, pasture, flower, or cloud, and the boys were the river gods Potamoi. The placing of the Atlas near the world ocean or ocean-sea, which was represented by the Atlantic Ocean, was unknown to a large extent to the ancient Greeks; it symbolizes that the Atlantic Ocean ruled by Oceanus, which was linked to Atlantis, when the inner body of water, the Mediterranean, was considered a sea and assigned for Poseidon to rule. Oceanus was the father of the rivers and streams, thus the progenitor of river gods. According to Critias, the Hellenic gods divided the land, and Poseidon was given the island of Atlantis. As Poseidon had five pairs of male twins with his lover with whom he built a fortress encircled with three rings of moat, the eldest of these twins, Atlas, inherited the whole main island (or continent) of Atlantis and also the mountain of his birth and the surrounding area as his fiefdom, which lie in Morocco, and are called Atlas, representing clear proof for this ancient denomination.

On the other side of the ocean, named presumably in Atlas's honor "the Atlantic," was situated the legendary ancestral home of the Nahua peoples, or Aztecs, called Aztlán, or *Aztlān* in Nahuatl (or Aztec) language, which means the land of the white saint. "Azteca" is the Nahuatl word meaning "the people from Aztlan." The similarity of the meaning of the denominations for other relevant Aztec gods mentioned in the etymological part is striking. More recently Aztlán is considered by some as the mythic ancestral home of the Mexican people, which in some sources is placed in the southwestern part of the United States. The mestizo activists of the Chicano movement even defined the nation of Aztlán, which they believe have legal and historical rights to the Mexican lands occupied by the United States during its 1846–1848 war with Mexico. While some legends describe Aztlán as a paradise, the Aubin Codex, one of the Aztec codices, says that the Aztecs were ruled by a tyrannical elite from which they fled, guided by their priest, a story similar to that of the fleeing of Jews from Egypt. The term Tlalocan, signifying the first paradise, could be a derivative of Atlaloc.

As Atlas or Atlaua or Atlatonan was representing at the same time significant Greek, Roman, and Aztec gods, the similar denomination could have been explained only by accepting the existence of Atlantis, which made a connection between these two worlds, exactly coming from the Atlantic or even from central America and conquering the Mediterranean, as accounted for by Plato.

The New Atlantis

Fictional societies are a common device employed by authors to illustrate political or social points, from George Orwell's *1984* and Aldous Huxley's *Brave New World* to Thomas More's *Utopia* and Francis Bacon's *New Atlantis*. Bacon's utopia of Bensalem was described as a scientific society where invention and experimentation were highly prized. Though religion plays a part in the story, Bensalem is unique as a New World civilization portrayed as more sophisticated than the Old World, more scientific rather than "savage," and the implication is that while Christianity may well be compatible with this sophistication, it is not a prerequisite for it—and therefore, outside the fictional realm of Bensalem, there may be reason and morality without appeal to supernatural sources.

Donnelly stated that all major ancient cultures are linked in one way or other with Atlantis's culture; others proposed that Atlantis was somehow related to the Mayan and Aztec culture.

Archaeological Evidences

Pyramids have so far only been found in Egypt and in Mesoamerica, and evidence was missing only from Europe to link the Atlantis hypothesis. The existence of the pyramids on both shores of the Atlantic near the same latitude is striking evidence of a common cultural legacy, since it represents a type of building unknown in other parts of the world. The recent but disputed discovery of ancient pyramids in Visoko, Bosnia, presumably dating to the period from the seventh to ninth millennia BCE, could explain the conquest of Europe by the Atlantidians, coming from the west, since they show some similarities with the pyramids of Teotihuacan and Tenochtitlan. Teotihuacan, meaning the "birthplace of the gods," is an impressive settlement and place of worship located just north of Mexico City, encompassing more American cultures and peoples including the Mayas, Aztecs, and possibly the Toltecs. Tenochtitlan was a 212,000 to 350,000 strong Nahua altepetl or city–state and was the Aztec capital by the time of the Spanish conquest in 1521. It is located on an island in a shallow lake, called Texcoco, in the valley of Mexico; its location resembles Atlantis, with water encircling it, pointing out to a probably preserved tradition of fortifications around the world. The style of building on islands cut off from the mainland by shallow straits of water is present worldwide and may have surfaced in Venice, in the region where the Bosnian pyramids are located.

Archaeologist and discoverer Semir Osmanagić (2005) maintains that these pyramids resemble the 1,800-year-old ones at Teotihuacanthe; the largest of them is bigger than the pyramid of Kh'ufu at Giza, Egypt, and they date to 12,000 BCE. This finding is disputed by Curtis Runnels, a specialist in the

prehistory of the Balkans, who claims that between 27,000 and 12,000 BCE the Balkans were locked in the last glacial maximum, a period of very cold and dry climate with glaciers in some of the mountain ranges, and that the only occupants were Upper Paleolithic hunter-gatherers who left behind open-air camp sites and traces of dwelling in caves (Runnels 2004).

In the archipelago of Bahamas, just off North Bimini Island, there is a submerged rock formation of large rectangular stones, called the Bimini Road, which has been claimed by some to be the evidence of the lost civilization of Atlantis. The road is localized in a depth of 6–30 meters, just off the coast of Bimini, and it consists of a row of huge rocks perfectly cut that continue on the sandy bottom for several hundreds meters. The Bimini Road is a 1,900-foot-long J-shaped formation of two parallel stone walls or docks in the shallow water off Bimini. The edifice was coined as an extensive underwater megalithic complex forming "an extensive pavement of rectangular and polygonal flat stones of varying size and thickness obviously shaped and accurately aligned to form a convincingly artifactual pattern" (Valentine 1976). The weight of only one 5-meter wide square or rectangular stone is about 15–25 tons, and some researchers accepted it as ruins from Atlantis. It was found that the massive stone blocks were not connected directly to the sea floor but were resting on the bedrock, and some claimed that the site was identical to numerous ancient man-made harbors discovered throughout the Mediterranean. Some were even held up by short columns. Similar large stones were found by Jacques Cousteau in the small island Dias of Crete in 1976. These walls could be the remnants of Poseidonia, the capital of Atlantis, or a festive road of a Maya town; however, some conclude that they were made ca. 10,000–7000 BCE, being older than the Mayan pyramids.

Andros is the largest island in the Bahamas, its name probably coming from the Greek word for "man." Near the island of Andros there is a large platform of massive, rectangular limestone blocks similar to those for the Bimini Road, which may represent a sunken temple 60 by 100 feet in size, square in form, having 3-foot thick walls, laid out in near-perfect 90 degree angles, extending 9 feet down into the sand, resembling stone stairways similar to Mayan sites. The Andros Platform closely resembles the Bimini Road, and one suggestion is that the structures built of beach rock could have been building platforms, but were probably breakwaters enclosing harbors with a quay nearby. The newly crystallized theory is that both structures were at harbors, probably elevated building platforms similar to some associated with 8,000-year-old or more ancient Mediterranean harbors. The location of the Andros Temple some 60 miles in the distance supports the idea that the deep channel running next to Andros was a shipping lane in ancient times. The Bimini Road and the Andros Platform would have been ideal stopping points for ships making the journey from the Atlantic to Cuba.

If the melting of Ice Age glaciers is responsible for the present depth of the ruins, they should date back roughly 7,500 years. The temple's size fits

extremely close to the Golden Section—used extensively by the Mayan engineers of Mexico and Central America in their buildings—this is clearly an edifice of importance built by a mathematically sophisticated civilization. The Andros Temple is partitioned into three separate rooms, identical in plan to the Temple of the Turtles at the ancient Mayan city of Uxmal in Yucatan. The edifice is a 100-foot-tall limestone outcrop, located at Nicholls Town. On Andros other edifices were found too near the temple site, like a pile of large rocks with a diameter of roughly 3 feet that continued 12 feet straight down with a wall of stones at the bottom where a 40-foot-long horizontal cave or tunnel started, suggesting that it was a lookout point over the ocean.

Sonar finds revealed rectangular formations lying in 100 feet of water off Bimini Island, several unusual stone formations approximately 20 miles out on the Great Bahaman Bank, and an underwater "mass" of fully dressed marble beams, an exquisite apex, columns, and numerous huge, rectangular flat slabs of white marble. In 2007 an underwater, partially intact wall, built from three to five layers of huge slabs of limestone, was discovered off a small, uninhabited island north off Andros. All of this evidence points toward a developed maritime civilization located in this region, which might be the remnants of Atlantis itself.

Islands reportedly covered in ruins and an ancient underwater harbor off the Yucatan are other signs of ancient civilizations. In the summer of 1927, at a Mayan site in British Honduras in Central America, a perfect skull made of one single transparent quartz crystal with a movable jaw was found in the jungle. The skull couldn't have been made by any modern method, not even by laser, and the lack of scratches suggests that no metallic tools were used. A carved 3.8-by-2.2-meter gravestone, presumably representing a pilot, was found in the temple of the signs of Palenque, Mexico, near the temple of Atlantis. The slab resting on a sarcophagus probably is the image of a pilot of an aircraft who is working in the cabin using different equipment among which levers, pistons, pipes, conductors, and cylinders can be distinguished, all these pointing to a very sophisticated earlier civilization.

Yonaguni, the westernmost island of Japan, lies 125 km (78 miles) from the east coast of Taiwan at the end of the Ryukyu Islands chain and is home of the Yonaguni monument, an impressive underwater structure. The monument seems to be a pyramid, 120 meters in length, 40 meters wide, and 20–25 meters high, claimed to be 8,000 years old. Parts of the monument are two closely spaced pillars that rise to within 8 feet of the surface, thought to be some 10,000 years old. The island would have been much larger when the ocean was 39 meters (100 feet) shallower at around 10,000 BCE. Submarine earthquakes are frequent, the last one from 1998 destroyed part of the island, which during the Ice Age composed a land bridge between Taiwan, Japan, and mainland Asia. As evidenced, the seabed contains what appears to be ruins of terrestrial flora, fauna, and stalactites, which form only on land. In Plato's account: "the Atlanteans

then built bridges northward from the mountain, making a route to the rest of the island. They dug a great canal to the sea, and alongside the bridges carved tunnels into the rings of rock so that ships could pass into the city around the mountain; they carved docks from the rock walls of the moats. Every passage to the city was guarded by gates and towers, and a wall surrounded each of the city's rings" (Plato, *Critias*).

The Possible Locations of Atlantis

Since the renewed interest for the location of Atlantis, there have been at least 30 locations proposed for Atlantis, but many proposed sites are not within the Atlantic or Mediterranean range. Many of the proposed sites share some of the characteristics of the Atlantis story, but none has been yet proven conclusively to be the true historical Atlantis. Though many location sites were proposed, stretching down from Sweden to Antarctica, the two most likely locations are the one in the Atlantic Ocean at the same latitude as the Mediterranean and the Mediterranean region that was home to Plato, the Egyptians, and to all the relevant classical authors.

Atlantis Located in the Atlantic Ocean

Evidently Plato's accounts place Atlantis in the Atlantic Ocean, which might have been named exactly because of that, for this we have to admit that the Pillars of Hercules had meant the Straits of Gibraltar. There are two great rocks on both the African and European sides, the European one is called the Rock of Gibraltar (Calpe in ancient times), which rises to 426 meters (1,396 feet) high, with impressive vertical walls unmatched and two big pillar like end rocks. Its modern name is deriving from the Arabic *Gib al Tarik*, meaning the rock of Tarik, or *Jebel Tariq*, meaning the mountain of Tariq. The straits were known as the Straits of Gades in antiquity (Pliny the Elder).

The location of Atlantis in the Atlantic is the most widespread and accepted version in popular culture and also which perpetuates the original Platonic ideal. The Pillars of Hercules means the Strait of Gibraltar as Pliny the Elder accounted for, known as such because Hercules the god "dug through" the once original mountain, leaving a 300-meter deep strait with cliffs aside from which the sea gained admission. The water indeed flows into the Mediterranean, causing up to 100-meter-high internal tidal waves since the inland sea evaporates a much larger quantity of water than the amount flowing into it from the Mediterranean basin's rivers.

The first candidate for Atlantis may be the Spartel Bank or Majuán Bank, which is a submerged former island located in the Straits of Gibraltar at 35°55'N 5°5'W near Cape Spartel. Its top currently lies 56 meters below the

surface, and it vanished under the surface around 12,000 years ago due to rising ocean levels from melting ice caps after the last glacial maximum. Spartel sunk slowly under the rising sea levels starting 20,000 years ago. During the most recent glacial maximum, the sea level was 135 meters below its current height, narrowing the Straits of Gibraltar and creating a small half-enclosed sea measuring 70 by 20 km between the Mediterranean and the Atlantic Ocean.

The Spartel Bank formed an archipelago in this small sea with the largest island measuring about 10 to 12 km across. With rising ocean levels, the island began to slowly shrink, but then at around 9400 BCE (11,400 years ago), there was an accelerated sea level rise of 4 meters per century, known as Meltwater Pulse 1A, which drowned the top of the main island. He proposes that the disappearance of this island was recorded in prehistoric Egyptian oral tradition for 5,000 years until it was written down by the first Egyptian scribes around 3000–4000 BCE, and subsequently inspired Plato to write a fictionalized version interpreted to illustrate his own principles.

Marine geologist Gutscher (2005) made a high-resolution map with sonar that indicated that the island was even smaller, meaning it would have been uninhabitable as long as 14,000 years ago, but the sediments gathered tell a different story, testifying to an earthquake of a magnitude 9 on the Richter scale, appearing to have rocked the region around the believed time of Atlantis's doom. By that time, the island would have been big enough to have been inhabited. He also found evidence of subsequent earthquakes and tsunamis—apparently every 2,000 years—that may have whittled the rest of the island away.

The candidates for the remnants of Atlantis are also the volcanic Canary Islands, which lie southwest of the Straits of Gibraltar but in close proximity to the Mediterranean Sea. The archipelago consists of seven major islands, one minor island, and several small islets. The name Islas Canarias is likely derived from the Latin term *Insula Canaria*, meaning island of the dogs, pointing at a dense population of an endemic breed of large and fierce dogs, which may have survived from the time of Atlantis, with the dogs as the main characteristic remembered most by the few ancient Romans who established contact with these islands by the sea. The connection to dogs is retained in these animals' depictions on the islands' coat-of-arms. The islands were known also to the Phoenicians and Greeks and are mentioned in a number of classical sources, like Pliny the Elder's seminal encyclopedic work of 37 volumes *Naturalis Historia* (Latin for "natural history") written at around 77 CE.

The islands could be one of the Fortunate Isles, or the Islands of the Blessed, of other classical writers, which were thought to lie to the west of the Straits of Gibraltar. These islands are referred to today as Macaronesia, along the ancient lines, continue to play the role of defining the prime meridian through the Middle Ages, and most probably have referred to the Azores, Madeira, the Canary Islands, and Cape Verde. They are all of volcanic origin and

are thought to be the product of several geologic hotspots. Ptolemy used these islands as the reference for the measurement of longitude, and Plutarch also locates them firmly in the Atlantic. The climate of the Macaronesian islands ranges from subtropical to tropical, and therefore these islands appear as a paradise, existing also in Greek and Celtic mythology. Philostratus said: "And they also say that the Islands of the Blessed are to be fixed by the limits of Libya where they rise towards the uninhabited promontory." In this geography Libya was considered to extend westward through Mauretania "as far as the mouth of the river Salex, some nine hundred stadia, and beyond that point a further distance which no one can compute, because when you have passed this river, Libya is a desert which no longer supports a population" (Philostratus 2009).

The Azores and other islands in this region of the Atlantic, like Madeira, are also candidates and fit into this theory. The theory holds that the Azores islands and in their extent Madera and the Canary Islands, are the emerged summits of a big undersea volcanic mountain range, where earthquakes and explosions often happen. The Canary Islands, Madeiras, and Cape Verde were identified by Spence (2003) as what was left of the easterly part of Atlantis when it broke up. They could have been the peaks of a submerged continent as the rocks from the "Telegraph square," situated at a depth of 2,800 meters, had been created by glassy basalt lava on the mainland and not by an undersea volcano, since they were solidified in the air. These rocks originated from an island that existed 15,000 years ago, created by a mainland volcano. Its sharp form suggests that it had not been corroded significantly yet by water. Later undersea findings confirmed that the same type of rocks existed in a huge area of that depth of the Atlantic. The initial finding from 1898 was estimated to be 15,000 years old, which confirms the age of destruction of Atlantis at 11,500 years ago or in approximately 9500 BCE.

The Caribbean Region

Geologists admit that between the Big and the Small Antilles there was land, which had sunken, reaching from Central America to the Big and the Small Antilles and up to the Bahamas, covering the Sargasso Sea, at its largest extent. This mass of land, or continent, was also surrounded by many small islands. This led some scientists to assert that Atlantis was located in the Caribbean and represented a highly evolved civilization that possessed ships and aircraft powered by a mysterious form of energy. The Sargasso Sea itself may be the proof of Plato's account as "also the ocean at that spot has now become impassable and unsearchable, being blocked up by the shoal mud which the island created as it settled down" (Plato, *Timaeus*).

The Bahamas are a large system of some seven hundred islands spread across the Bahaman shelf. During the last Ice Age, nearly 12,000 years ago,

much of the Bahaman shelf was above sea level. Rainfall during that period formed erosional-type gullies or small canyons as rainwater flowed off the shelf into the tongue of the ocean. The Bahaman shelf may be the flooded continent Atlantis, and the tongue of the ocean may have been the channel leading to the city of Atlantis. Plato describes Atlantis as being long and narrow from east to west, and that it could be crossed on foot in 18 days and nights. Afterward, by sailing from some other islands, one could reach the continent, which was situated just opposite.

For instance, the biggest part of it was once one island with its middle submerged. Its rims were formed by what are now the Crooked and Acklins Islands. The New Providence and Eleuthera Islands show a similar picture, while the greatest part of Andros Island is also submerged slightly, but it is still of considerable size: 167 km long and 64 km wide at its widest point. The same holds for the Grand Bahama Island, which is still 96 miles (154 km) long and 17 miles (27 km) wide at its widest point, as a clear remnant of a sunken continent. Interestingly the island of Andros has its identical named counterpart in the upper Cyclades, north of Athens in the Aegean. The water-filled caves of the otherwise flat island, like that of the Guardian Blue Hole, contain stalagmites at 100–200 feet below sea level, proving that the whole island had submerged substantially.

Cuba is also a possible location for Atlantis. Physicists determined that ca. 10,500 BCE a fragmenting comet broke into two large pieces before it struck the Atlantic (known as the Carolina Bays event). As a consequence, millions of smaller fragments hit the eastern coast of America as well as the islands in the Gulf of Mexico and the Caribbean. Massive flash fires instantaneously burned through the islands and minutes later a series of massive tidal waves swept the islands clean of structures and human life. Thus the maritime culture of Atlantis would have extended from Spartel Island at Gibraltar, through the Azores, to the much-larger Bahama Islands and on to Cuba. In 2001 at the extreme western tip of Cuba at Zapata, ruins of buildings were discovered and located through side-scan sonar at a depth of 2,200 feet (700 meters). The location may conceal the capital city of Atlantis.

Atlantis Located in or Near the Mediterranean Sea

Most of the historically proposed locations are in or near the Mediterranean Sea. The most popular is Santorini or Thera, where the eruption of the volcano located under the island occurred around the 17th or 16th century BCE, causing a massive tsunami that probably hit the shores of Crete and devastated the Minoan civilization developed on the island.

The elements of Plato's story correspond to this location on many accounts: the possible existence of circular mots, the size of the main city–island roughly

corresponds with the one given in Critias: "the central island itself was of a stade's breadth or 607 feet long" (Plato, *Critias*), and the island is made of white, red, and black stones. The earthquake followed the eruption of the volcano, which blasted the center of the island, with the surroundings being sunken immediately, so it could have happened overnight. The actual island (or parts of the island) are the remnants of the outer rim of the original island and therefore a popular and well-publicized tourist destination. However, Plato didn't say anything about a volcanic eruption, but only about an earthquake and the following tsunami, and also the size of the island is far from being a continent or a large island of hundreds of kilometers or miles in width. This size rather fits a large island such as Crete, Sardinia, Sicily, or Cyprus. Angelos Galanopoulos and Edward Bacon argued that the time scale has been distorted by an error in translation, probably from Egyptian into Greek, which produced "thousands" instead of "hundreds"; this same error would rescale Plato's kingdom of Atlantis to the size of Crete, while leaving the city the size of the crater on Thera, and dating to 900 years before Solon, which would be the 15th century BCE (Galanopoulos and Bacon 1969).

Crete was home to the advanced Minoan civilization, a naval power that presumably was peaceful, based on gender equality, with males being represented as having red skin while the females had white. This may correspond to the description of Atlantis castle as having "walls constructed of red, white and black rock quarried from the moats, and covered with brass, tin and orichalcum" (Plato, *Critias*). The latter means mountain copper, which covered the interior of the whole citadel, which "flashed with the red light of orichalcum." This material was the legendary precious metal, an alloy, or simply red copper. The outposts of the Minoans were scattered on many neighboring islands, including Santorini, situated at around 70 miles (110 km) distant. The existence of this Aegean archipelago is evidence of a once larger island and many islets. The Minoan civilization on the northern and eastern shores of Crete was destroyed by the tsunami following the Santorini eruption around 1600 BCE, which destroyed first Akrotiri, the Minoan town on Thera. The Thera volcano periodically experienced violent eruptions, then eventually collapsed into a roughly circular seawater-filled caldera, with numerous small islands forming the rim of the caldera. The caldera could slowly refill with magma, building a new volcano, which erupted and then collapsed in an ongoing cyclical process, as it could have been observed in the case of Krakatoa in Indonesia. The physical consequences of the eruption—a tsunami from 35–150 meters (115–492 feet) high swept across the region, reaching even 250 meters (820 feet) on the island of Anafi, situated 27 km (17 miles) to the east, weakened the Minoan civilization, and allowed the Greek Mycenaeans to conquer it.

Troy (Greek: Τροία, or *Ilion*, Turkish: *Truva*) is a legendary city whose probable location is the archaeological site in Hisarlik in Anatolia. Troy is now

5 km from the seacoast, southwest of the Dardanelles under Mount Ida, but in 3000 BCE, a natural bay reached up to the city's walls. The ancient mouths of alleged Scamander River had since been filled with alluvial material, and recent findings confirm the accuracy of the Homeric geography of Troy. The capital of Atlantis had a castle. "Fifty stadia (10 kilometers; 6 miles) from the coast was a mountain that was low on all sides . . . broke it off all round about" (Plato, *Critias*). The layers of the ruins at Hisarlik date back to 3000–2600 BCE, so if the Platonic account is downscaled by using hundreds instead of thousands, it is theoretically possible to be the site of the Atlantic capital, so in this case Anatolia or today's Turkey would have been Atlantis. A small minority of contemporary writers argue that Homeric Troy was not in Anatolia, but located elsewhere—in England, Croatia, Transylvania, or Scandinavia—although these theories have not been accepted by the mainstream. Troy's legends were transmitted to the ancient Greeks through Homer's *Iliad* and also mentioned in the *Odyssey* and in the *Aeneid*, by the Roman Vergilius Maro. According to Greek mythology, the Trojans were the citizens of the ancient municipality of Troy, as if it were part of the Greek culture of city–states, whose people understood Greek, but this is uncertain. In fact, anyone from its jurisdiction, which was mainly the Troad region, might be called a "Trojan." Troy was known for its riches gained from international trade and metallurgy and for its massive defensive walls. According to the Greeks, the parents of Dardanus, the founder of Troy, and of Dardania, were gods. The kingdom passed to his grandson Tros, who called the people Trojans, and Ilus, son of Tros, founded the city of Ilium (Troy). Poseidon and Apollo built the walls and fortifications around Troy for Laomedon, son of Ilus the Younger, but when Laomedon refused to pay, Poseidon flooded the land and demanded sacrifice to a sea monster. As a consequence, pestilence followed and the sea monster snatched away the people of the plain.

This doom, which might correspond to the flooding of Atlantis, was prophesied by Cassandra, who was granted this gift by the god Apollo, in order to return his love. When she failed to do this, Apollo placed a curse on her so that no one would ever believe her predictions. Cassandra or Alexandra, meaning "the she who entangles men," was the daughter of King Priam and had told about the Trojan horse, but according to some, she was considered mad and incarcerated, therefore driving her truly insane in the end. In moderate versions, she was not considered mad and her prophecy was simply misunderstood. The Heracleid dynasty's founding myth asserts that one generation before the Trojan War, Heracles captured Troy and killed Laomedon and his sons, excepting the young Priam. During Priam's reign, the Mycenaean Greeks invaded and captured Troy in the legendary Trojan War fought between 1193–1183 BCE.

Antcomah, or "a place to meet in wartime" in Greek, is mentioned in the *Iliad* and refers to two possible sites: a Greek village in Kos (or Turkish: İstanköy) Island in the Aegean Sea, called Αντιμάχεια, or a now abandoned

ancient Greek town near Trabzon, also in Turkey, where an archaic Greek dialect persisted until recently since the seventh century BCE. According to the legend it was a very rich city before the Bosporus appeared, but lost its importance as the seashores had withdrawn. Supposedly the site is still very rich in iron and the small lake of about 30 meters wide hides a fortune under the water.

In the Black Sea area, the straits of Bosporus are also a possible location for Atlantis, being the site of a submerged part of the continental shelf. The exact cause for the formation of the Bosporus remains a subject of vigorous debate among geologists. Thousands of years ago, the Black Sea became disconnected from the Aegean Sea. One recent theory contends that the Bosporus was formed about 5600 BCE, when the rising waters of the Mediterranean/Sea of Marmara breached through to the Black Sea, which at that time was a low-lying body of fresh water. Others argue that it was the result of massive flooding of the northern shores of the Black Sea, which is now shallow and uninhabited, but could have destroyed a developed civilization such as Atlantis. Many have argued that this flood may be the historic basis for the flood stories found in the Epic of Gilgamesh and in the Bible. An opposite direction flood, caused by the inflowing rivers and heavy rainfall, might have led to an inflow into the Sea of Marmara ca. 7000 or 8000 BCE and destroyed such city–states as Troy.

Another location in the Black Sea region is the nearby Sea of Azov or Azak, meaning "low" in Turkish, being the shallowest sea in the world, with an average depth of 5–13 meters (43 feet) and a maximum depth of 15.3 meters (50.2 feet), and only 1 meter in the bay of Taganrog, which is fed by four rivers. The large bay, 140 km in length and 31 km in width at the mouth of the Don, also abounds in sandy spits that partly enclose shallow bays. The bay contains the Sandy Isles (Песчаные острова), and the sea itself is also the freshest, where salinity can be 30 to 40 times lower than in other seas and where enormous quantities of silt are deposited by the inflowing rivers Don and Kurban. Such a body of land of this size (343 by 231 km) with a surface of 37,605 square km resembles the remnants of Atlantis, which might have been flooded by the inflowing sea when the strait of Kerch has opened. In Plato's account, Atlantis was "a haven having a narrow entrance; but that yonder is a real ocean, and the land surrounding it may most rightly be called, in the fullest and truest sense, a continent" (Plato, *Critias*).

The Black Sea deluge theory states that the sea of Azov, as well as all the large northern shallow coastal bay areas, was flooded by the ocean coming through the straits of Bosporus, an event that occurred ca. 5600 BCE. During the Ice Age, the Black and the Caspian seas became vast freshwater lakes, due to the inflow of great rivers (Danube, Volga, Don, Dniester). As the glaciers retreated, rivers emptying into the Black Sea reduced their volume and changed their outflow in the direction of the Baltic and North seas, while increased temperatures

lowered water levels through evaporation in the Black Sea, and meltwater increased it in the Mediterranean. Then, the rising Mediterranean finally spilled over a rocky sill at the Bosporus, flooding around 60,000 square miles (155,000 square km) of land. The theory is disputed but also confirmed by marine archaeologist R. Ballard's series of expeditions, during which he identified seemingly ancient shorelines, drowned river valleys, tool-worked timbers, and man-made structures in roughly 300 feet (100 meters) of water, off the southern coast of the Black Sea (Ballard 2001). Radiocarbon dating of freshwater mollusk remains indicated an age of about 7,000 years, which may not be the "accepted" timeline for Atlantis's disappearance, but there are strong arguments for the theory. The theory was confirmed by a cooperative European Union project called ASSEMBLAGE and completed by the Noah Project led by the Bulgarian Institute of Oceanology.

Accounts of a lost ancient civilization are to be found also on Malta (Melita or "honey" in Greek, Maleth or "heaven" in Phoenician), a group of seven islands with the island of Malta in the center. Malta's location is also perfect for conquering Libya and the Tyrrhenian Sea region. The Maltese Islands were first settled ca. 5000 BCE by Stone Age farmers from the larger nearby island of Sicily, but these dates may be erroneous. At ca. 3500 BCE these people built the oldest surviving freestanding structures and oldest surviving religious structures in the world, in the form of the megalithic Ġgantija, Ħaġar Qim, and Mnajdra temples. Ġgantija is believed to be built by giants; its walls in places are still standing to a height of 7 meters. The location of the prehistoric megalithic temple complex Mnajdra corresponds with Plato's account, being close to the shore in the south of Malta, surrounded by a circular wall and situated very close (approximately 500 meters) to the Ħaġar Qim megalithic structure; thus together they may form the two parts of the capital city. Mnajdra was build around the third millennium BCE, and therefore it is one of the oldest freestanding buildings in the world, predating the Egyptian pyramids and Stonehenge. It is made of white limestone. Traces of an ancient and suddenly disappeared civilization are also the Hal Saflieni Hypogeum ("underground" in Greek), an enormous subterranean necropolis of more than 500 square meters, the only prehistoric underground temple devoted to worship and burial, where the bones of over 7,000 people have been found. The system of caves, passages, and cubicles perhaps had a steep drop opening to the sea according to mysterious accounts and disappearances. White limestone is accompanied by red ochre, which is the building material for spiral mosaics and wall designs, and the cult of the bull can be also discovered by admiring the outline of a bull and the shape of a black hand.

Sicily is the largest island in the Mediterranean and with its smaller surrounding islands was a separate kingdom and civilization for thousands of years, its first inhabitants being the Sicani, who, according to Thucydides, arrived from the Iberian Peninsula, perhaps from Catalonia, so this fact fulfills the Atlantis criteria. Since important historical evidence has been discovered in

The megalithic temple complex of Mnajdra on the island of Malta is one of the oldest and most striking examples of neolithic construction. (Shutterstock)

the form of cave drawings by the Sicani, dated to the end of the Pleistocene epoch, around 8000 BCE, this can be counted as a proof of the existence of more ancient civilizations, founded by Atlantidians. The painter outlived the destruction of the core area in presumably 9000 BCE. The island is dominated by volcanoes, Etna being the most prominent, from the triangle composed of Europe, northern Africa, and western Asia. In the region there are volcanic islands such as Stromboli that might have been sunken in prehistoric times.

Other Locations

In and around the Mediterranean, up to the North Sea and down to the Indian Ocean and Pacific, other islands or once existed land regions were proposed, where traces of megalithic cultures can also be found. These include Sardinia, Cyprus, and the British Archipelago, the latter being disconnected from the European mainland by the meltwater of the end period of the last Ice Age, in a period approximately corresponding to Plato's account, the islands being home to Neolithic civilizations. Smaller islands or submerged areas, such as the Sole Bank or the islands of Scilly (not to be confused with Sicily), may also be

considered. The Sole Bank is a relatively shallow area in the Atlantic, 100 miles off England's Land's End. The area surrounding Land's End abounds in possible locations, such as midway between it and the Isles of Scilly, which is the supposed location of the mythical lost land of Lyonesse, referred to in the legends of Brythonic peoples.

The islands of Scilly, inhabited since the Stone Age, were known in the antiquity to the Phoenicians as the Cassiterides (Tin Isles) and indirectly to the Greeks. It is likely that until relatively recently the isles were much larger, with many of them joined into one island, and that the land has subsided, as an ancient description from the time of the Roman Empire describes it as "Scillonia insula" in the singular. At certain low tides the sea becomes shallow enough for people to walk between some of the islands. Ancient field walls are visible below the high tide line off some of the islands, and remains of prehistoric farms have been found on a small rocky skerry far too small for farming. Seven of the islands of Scilly are or have been inhabited until recently and another 49 belong to the archipelago, where subtropical vegetation is present on the southern shores of those islands, which have a protective morphology against the Nordic winds. Proclus in his fifth-century commentary on Timeus, wrote that Atlantis "was comprised of seven islands in that sea in their time, and also three others of enormous size" (this might correspond to Britain, Ireland and the Dogger Bank) (Proclus 1820).

On the other side of Britain, about 100 km off the coast the Dogger Bank, is a moraine formed at the southern extent of glaciation during the last Ice Age which seems to be a good candidate. The bank was land during the last Ice Age, part of a larger landmass known as Doggerland, which connected Britain to the European mainland. As evidence, a large amount of great mammal bones and some Paleolithic hunting artifacts were found on it. Its maximum dimensions were about 260 km (160 miles) from north to south and 95 km (60 miles) from east to west, covering a size of approximately 17,600 square km (6,800 square miles). Therefore, it corresponds to a certain extent to the Egyptian description for Atlantis as cited in Critias: about 555 km (345 miles) long and about 370 km (230 miles) wide. The water depth ranges from 15–36 meters, and 23 kilometers beneath the bank was the epicenter of the largest earthquake ever recorded in the United Kingdom, which happened in 1931 and was felt in all countries around the North Sea. The German island of Helgoland and mainland based cities or states such as Tartessos, at the western end of Andalusia and at the mouth of the Quadalquivir, which had disappeared suddenly in the sixth century BCE, are also mentioned as candidates. This may correspond to the mouth of the river once forming the lake of Ligustinus. The location of the Aztec capital may fit Plato's story as well: "Poseidon carved the mountain where his love dwelt into a palace and enclosed it with three circular moats of increasing width, varying from one to three stadia and separated by rings of land proportional in size (Critias 116 BCE)" (Plato, *Critias*).

Other more exotic locations include Indonesia, Malaysia, or both, especially the Sundaland, which is a partially sunken biogeographic region located on the Sunda shelf and forming the boundary of the Indomalaya and Australasia ecozones. Genetic research carried out by Oppenheimer (1999) indicates that the remnants of Sundaland were likely populated 50,000 years ago, contrary to a previous hypothesis that population migrated there during the last Ice Age from Taiwan. This fits the Egyptian reference for an island–continent larger than ancient Libya and Asia Minor combined. He also claims that it was the world's leader in the Neolithic revolution of starting agriculture, as early as 24,000 ago, more than 10,000 years older than Egyptian civilization. Before and especially during the gradual flooding of their lowland, after the last Ice Age which was the biggest migration trigger in human history, the Sundalanders spread out to neighboring lands: the Asian mainland including China, India, and Mesopotamia, the island world from Madagascar to the Philippines, and New Guinea, where they later colonized Polynesia. Another lost continent is located in the same region, off the southern shores of India's Tamilnadu, which encompassed Sri Lanka and the Maldives as well, which was called Kumari Kandam. Recent linguistic research shows a similarity between the Indo-European and Austronesian languages such as Malay, Tagalog, or Maori, suggesting that the Aryan invasion into India was taking place from the southeast rather than from the northwest. Location suggestions for Antarctica remain unlikely however, since geologic evidence shows that it has never been habitable in the past 100,000 years. Australia as well as many islands may fit for a location somehow in the sense as the Egyptians described Atlantis as an island comprising mostly mountains in the northern portions and along the shore, and encompassing a great plain of an oblong shape in the south.

References and Further Reading

Antonopoulos, J. "The Great Minoan Eruption of Thera Volcano and the Ensuing Tsunami in the Greek Archipelago." *Natural Hazards* 5 (1992).

Ballard, Robert, et al. "Deepwater Archaeology of the Black Sea: The 2000 Season at Sinop, Turkey," *American Journal of Archeology* 105:4 (2001), 607–623.

Berlitz, Charles. *Atlantis: The Eighth Continent*. New York: Putnam, 1984.

Braswell, Geoffrey E. "Introduction: Reinterpreting Early Classic Interaction." In Geoffrey E. Braswell, ed., *The Maya and Teotihuacan: Reinterpreting Early Classic Interaction*. Austin: University of Texas Press, 2003.

Cayce, Edgar Evans. *Atlantis*, New York: Hawthorne, 1968.

De Camp, L. S. *Lost Continents: The Atlantis Theme in History, Science, and Literature*. New York: Gnome Press, 1954

Donnelly, I. *Atlantis: The Antediluvian World*. New York: Harper, 1882.

Galanopoulos, Angelos G. and Edward Bacon. *Atlantis: The truth behind the legend*. New York: Thomas Nelson & Sons, 1969.

Gordon, J. S. *The Rise and Fall of Atlantis: And the Mysterious Origins of Human Civilization*. London: Watkins, 2008.

Griffiths, J. P. "Atlantis and Egypt." *Historia* 34 (1985).

Gutscher, Marc-André. "Destruction of Atlantis by a Great Earthquake and Tsunami? A Geological Analysis of the Spartel Bank Hypothesis." *Geology* 33, 8 (2005).

Oppenheimer, Stephen. *Eden in the East*. Phoenix: Orion, 1999.

Osmanagic, Semir. *Svijet Maja* (*The Mayan World*). Piscataway, NJ: Gorgias Press, Euphrates, 2005.

Philostratus. *The Life of Apollonius of Tyana*. ed. by Mario Lampić, 2009. Available at http://www.scribd.com/doc/12697046/Philostratus-The-Life-of-Apollonius-of-Tyana. Accessed July 12, 2010.

Plato. *Critias*, translated by Benjamin Jowett; alternative version with commentary.

Plato. *Timaeus*, translated by Benjamin Jowett; alternative version with commentary.

Proclus. *The Commentaries of Proclus on the Timaeus of Plato in five books; containing a treasury of Pythagoric and Platonic physiology*. trans. by Thomas Taylor. London: Thomas Taylor, 1820.

Ryan, William, and Walter Pitman. *Noah's Flood: The New Scientific Discoveries About the Event that Changed History*. New York: Touchstone, 1997.

Runnels, Curtis and Priscilla Murray. *Greece Before History: An Archaeological Companion and Guide*. Stanford, CA: Stanford University Press, 2004.

Spence, L. *The History of Atlantis*. Mineola, NY: Dover, 2003.

Valentine, J. Manson, "Underwater Archaeology in the Bahamas," *Explorers Journal* 1976.

Valentine, J. Manson. "Archaeological Enigmas of Florida and the Western Bahamas." *Muse News* 1, 2 (June 1969).

Weaver, Muriel Porter. *The Aztecs, Maya, and Their Predecessors: Archaeology of Mesoamerica*, 3rd ed. San Diego: Academic Press, 1993.

CON

Scholars, novelists, and romantics exist who hope to find some tangible proof of the existence of Atlantis, but Plato's story of Atlantis is a fable inspired by a significant number of cultures and events in Greek memory. Critias, the

student of Socrates, spins a yarn aimed to support Socrates's view of the perfect state. In doing so, his Atlantis is as mighty and aggressive as the Minoans, the Persians, and perhaps, the Athenian Empire. His Athens—the Athens of old, that is—is the ideal state, a just state. Athens earns glory for stopping the aggressive Atlantis by organizing all of Greece and deploying her extraordinary army of warrior guardians, an army that reminds us much of the Spartans. We have all forgotten this epic event and the great state Athens once was. The philosophers will remind us, they say, because only they have the ability to relate such a story properly. Plato's aim is to teach us of the perfect republic. If we learn it well, perhaps we can restore it to its former glory. So while many may read the story of Atlantis and go searching for her remains, the legendary island is merely a foil, a ruse for the real aim of the exercise: the glorification and rehabilitation of Athens at a time when the classical era was coming to an end.

The Source

Plato records the so-called history of Atlantis in two of his dialogues: the *Timaeus* and the *Critias*, and for these this section will rely on the Benjamin Jowett (1952) translation found in the Robert Maynard Hutchins edition of *The Dialogues of Plato*. We should note that Plato is the *only* source for the story of Atlantis, and he mentions the story in *only* these two dialogues. Both dialogues represent an attempt by Socrates's students, Timaeus and Critias, to entertain their teacher with "a feast of discourse." In the tradition of the Greek symposium, the students hope to impress their teacher with a fine story and finer arguments to show their mastery of the material and methods that Socrates has taught them in an earlier discussion. There also is a competitive bent to the students' performances: both wish to impress their master and both wish to outdo one another.

Socrates's subject in the previous meeting was the nature of the perfect state; his students were to present stories of how their own state, Athens, had waged war in a "becoming manner," showing "greatness" and "magnanimity" worthy of their city–state's tradition and education. The subject of Atlantis is mentioned briefly by Critias as an introduction to the creation epic that dominates the *Timaeus*. Here Critias explains how the story of Atlantis came from an Egyptian priest who claimed his culture was superior and older than that of the young Greeks. The creation account presented by Timaeus brings in references to other creation epics known to the Greeks, adding Socratic logic for support, along with a lesson on human anatomy. In this, Timaeus aims to present a "scientific" account of the beginning of all things to counteract the charges of Athens's immaturity made by Critias's Egyptian priest. There are undercurrents to the creation epic that challenge the notion that the Greeks were as immature

as the Egyptian priest maintained. There is also a theme of duality: of the perfect and the image of the perfect; of the heavens and the image of the heavens found in creation. This duality, this image of reality as compared to perfection, is directly related to the utopian Atlantis and her comparison to Athens's and Socrates's ideal state.

The dialogue attributed to Critias is wholly concerned with the topic of Atlantis. When Critias discusses Atlantis in both dialogues, it is worth noting that he is worried about his memory. He first heard the story as a child. He begged his grandfather to tell it to him again and again. There was once a written version, but that disappeared long ago. When he begins the story, he calls on the goddess Mnemosyne, "memory," to aid his recollection. He has rehearsed the story for his friends before relating it to Socrates so that he can get it right. He is astonished to note, as he remembers more of the story, how much the story of Atlantis reminds him of the lessons learned from Socrates. How much indeed! A further fault of Critias's story is not the survival of his memory but the survival of our manuscript: the *Critias* is fragmentary, breaking off just as Zeus enters the narrative to deliver a judgment against Atlantis.

Finally, the creation epic and story of Atlantis presented in the *Timaeus* and *Critias* must be told by philosophers, because according to Plato, philosophers are superior to poets and sophists in education and in knowledge of a city–state. Poets only imitate, sophists are too itinerate in their search for clients who will pay for their services to have full understanding of how a community deals with its day-to-day problems, let alone statecraft. Thus only Socratic philosophers, by their nature, their education, and practical political experience, are qualified to relate such epic stuff.

The Story

Drawing from both *Dialogues*, the story is as follows, as related by Sarah B. Pomeroy and colleagues in *Ancient Greece: A Political Social and Cultural History*. Critias has it on good authority that his great-grandfather, Dropides, heard the story from his colleague and friend, the legendary lawgiver Solon, that the island civilization of Atlantis was an "actual fact" not "mere legend." Solon, noted for his far-flung travels, heard the story from an Egyptian priest he met at the city of Sais in the Nile Delta. Solon had a written Egyptian version he had translated into Greek. Critias saw the manuscript as a child. His grandfather— also named Critias—once possessed the manuscript, but it was lost long ago.

The Egyptian priest was not surprised that Solon did not know the story, because the Athenians were "young" in their understanding of world history, unlike the Egyptians. The Athenians gained and lost the ability to write and record history many times over as a result of natural and military disasters. Additionally most of the players involved in their ancient past died in some of the

great events. Even the oral tradition was lost. The priest would now relate the story of Atlantis and Athens for several reasons. First Sais and Athens shared a patron goddess in Athena. Both cities had traditions of wisdom and education and both had well-organized states. Finally the priest now had an Athenian of great wisdom—namely Solon—with whom he could share this history.

The Egyptian priest related how 9,000 years before the time of Solon (who flourished in the 500s BCE) there existed a great island just west of the Pillars of Heracles (the Straits of Gibraltar) in the Atlantic Ocean. The island was so large that it was larger than Libya and Asia combined. (Libya was a portion of North Africa, and the Greeks deemed modern day Turkey as "Asia.) It was, in effect, an island continent. When Zeus, Poseidon, and Hades divided creation among themselves, Poseidon won Atlantis. He created the island civilization by fathering children with Cleito, a mortal woman, orphaned on the island. To protect the fledgling community, Poseidon made the center of the island an island itself, surrounding it with rivers. He and Cleito were the parents of five sets of twin boys. The first born twins would become kings, with the eldest of that pair having pride of place and giving his name, Atlas, to the civilization. The remaining sons were princes and became the ruling, hereditary aristocracy. Their descendants ruled the island civilization with a divine law code provided by Poseidon. They came together as a council every fifth and sixth year to render judgments and deal with problems of the island and between one another.

Atlantis had great success. With two growing seasons a year, the island enjoyed abundant natural resources. As great engineers, the people of Atlantis fashioned canals for communication, as well as aqueducts, roads, great palaces, and temples. The military capabilities of Atlantis were no less impressive. Atlantis maintained a large fleet and a large land army. The leaders of each region provided chariots, archers, slingers, and other military personnel when called upon. A great wall surrounded the island for protection.

The population grew under the patronage of Poseidon and his many sons. Atlantis came to control the western Mediterranean, including the western part of North Africa and modern-day Spain and France. The power, wealth, and stability of Atlantis were so great that the ambition of the king and princes grew as well. They planned to take over all the civilizations ringing the Mediterranean and thus waged war on Europe and Asia.

By comparison, Athens also enjoyed a golden age. The soil, Critias explained, was much more productive than in his own day. The climate was fine and education flourished. The political order divided Athens's population into classes according to nature and ability. There were herdsmen and artisans and, most important, the warrior guardians. Much like the guardians advocated in Plato's *Republic*, Athens's guardians maintained strength of mind and body through education and exercise. They lived in common, eschewing gold and silver and other luxuries. Their only aim in life was the defense and stability of their country.

When Atlantis became aggressive, Greece, under the leadership of Athens and her 20,000 guardians, defeated the mighty Atlantic powerhouse. Zeus followed the military defeat with a great earthquake and flood. The king of the gods had grown weary of the leaders of Atlantis, who had lost their brilliance. Through succeeding generations, that "divine portion" of them had begun "to fade away," as their baser, "human nature got the upper hand."

Atlantis Story as Rhetorical Exercise

With all due respect to those scholars, novelists, and romantics who hope to find some tangible proof of the existence of Atlantis, this pretty story of the greatness of Athens told by Socratic philosophers hoping to prove the superiority of their nature and education as the best of all thinkers and artists is simply a rhetorical exercise. The whole of the story includes references to a recent history more than 9,000 years before Solon's legendary excursion to Egypt. Through a close reading of the *Dialogues*, one can see connections to several great civilizations known to the Greeks: the Minoan civilization of the Aegean, the Persian Empire—an entity so vast that it touched three continents, the Spartans, and the Athenians themselves.

Essentially the Peloponnesian War ended the classical era of fifth-century Athens. Sparta won the war with the help of the Persian Empire and the poor leadership of Athens by several of Socrates's former pupils. The story of Atlantis is an effort to recapture some of the former glory of Athens in her time of defeat and confusion. Through the story, Plato can also favorably compare the antiquity of the Greeks to the older civilization of Egypt. Further, the defeat of a vast empire through the effective leadership of Athens is a pretty "pick me up" at a time of devastation and loss, when Athens's leadership was questionable. Atlantis was a well-organized state, richly endowed. She was a worthy adversary for the Athens of long ago: the greater the adversary, the more glorious the victory. After all, so effectively was Atlantis defeated by both man and god that she literally disappeared from the face of the earth. Finally, the political organization of Athens of old mirrors that of the utopian Republic of Plato. If Atlantis actually existed, and the brilliant Athens of 9,000 years ago actually existed, might Plato's *Republic* be more than a utopia one day?

Minoan Crete and Thera

Inspiration for Plato's Atlantis can be found in Minoan civilization. Much of Athenian mythmaking includes references to the older Minoan culture of the Aegean. Centered on the island of Crete, according to Arthur J. Evans (1964) in *The Palace of Minos at Knossos*, legend maintained what archaeology supports, namely that Crete was a large, wealthy, and productive island that may have
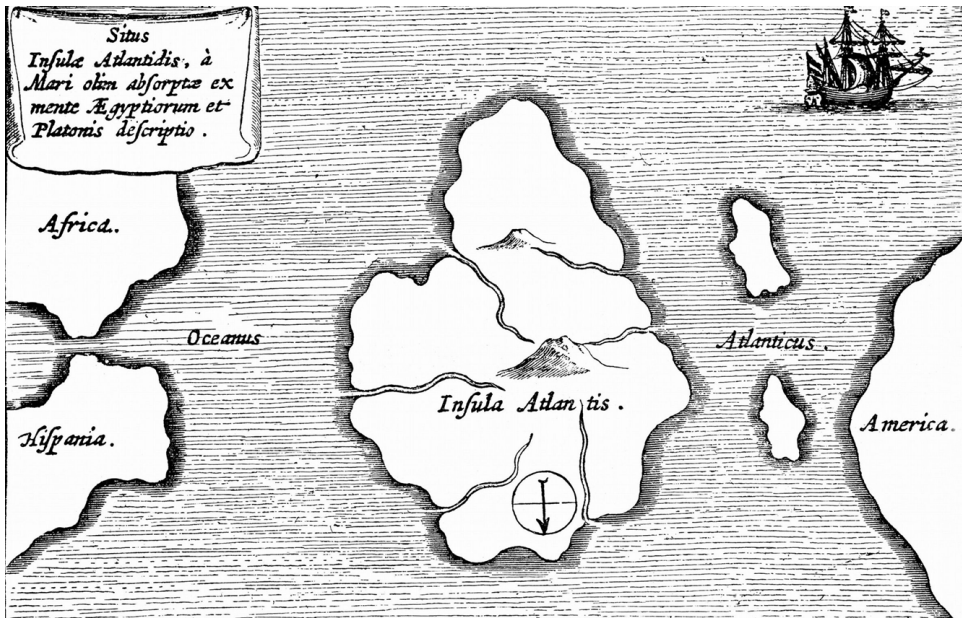
A rendition of the legendary island of Atlantis. Atlantis was first described by Plato and was said to lie just beyond the Pillars of Hercules (Gibraltar and Mount Hacho). This engraving of Atlantis represents the way that German Jesuit scientist Athanasius Kircher (1602–1680) envisioned the island. (Jupiterimages)

controlled a number of Aegean islands in a thalassocracy. Artifacts from Crete show a wealthy civilization with a palace economy. The Minoans traded with Africa, Asia, and Europe. Like the island of Atlantis, Crete had more than one growing season a year. The island power, ruled by a legendary king named Minos, was a special favorite of Poseidon, much like the island of Atlantis.

The well-known story of Theseus and the Minotaur, as related in Walter Burkett's *Greek Religion* (1985), shows a great connection in lore between the mainland Greeks and the island Minoans. In that story, a young prince takes up the challenge to end Minoan rule of Athens and other Greeks. According to the myth, King Minos, leader of the great sea empire of Crete, demanded a grisly tribute from his subjects: human sacrifice. The Minotaur, half man–half bull, ate only human flesh. The monster was the unlikely offspring of a prized bull given to Minos by Poseidon, and Pasiphae, the wife of Minos. Poor Minos not only had the problem of an adulterous wife with a penchant for bestiality, but also the monstrous offspring of that unholy coupling with a sacred bull of Poseidon. What to do? Minos placed the Minotaur in a great labyrinth and tossed in human sacrifices from among his subjects. The Athenian hero, Theseus, armed with a sword and ball of twine, killed the Minotaur, thus ending Minos's terrible reign.

The island of Thera (modern Santorini) to the north of Crete seems to have been a part of that culture, given analysis of the archaeological remains first

begun there by Spyridon Marinatos (1939). In the 1500s BCE, Thera suffered a devastating natural disaster. Thera (modern day Santorini) was an active volcano. In the 16th century BCE, that volcano imploded with such force that much of the island was destroyed. The extraordinary underwater caldera, created over successive eruptions, attracted many ancient tourists, as it still attracts tourists today. Scholars, like Marinatos in "The Volcanic Destruction of Minoan Crete," have speculated that the massive tsunami that would have followed such an implosion might have crippled, if not destroyed, a fleet of Crete. Regardless of the speculation regarding the relationship of Thera's implosion and the decline of Minoan civilization, the archaeological record shows that Greek writing overtook Minoan writing on Crete as early as 1400 BCE.

Plato writes that after Athens defeated Atlantis, violent earthquakes and floods followed, and "in a single day and night of misfortune" all the warriors of Athens sank into the earth and the island of Atlantis disappeared into the depths of the sea. Thus a combination of warfare and natural disaster ended the power of Atlantis and wiped out Athens's memory of the engagement. The myth of Theseus suggests that an Athenian hero put an end to a monstrous Minoan hegemony. The archaeological record demonstrates that Crete was a wealthy state, engaged in extensive trade. Greek writing replaced Minoan writing at one point, indicating that Greek culture came to dominate the island in the Bronze Age, confirming some aspect of the Theseus myth. The natural disaster that obliterated a significant portion of Thera could easily have inspired the idea of a natural disaster that sank a hubristic Atlantis. Minoan civilization and the devastation of Thera were known events in the fourth century BCE when Plato wrote.

Persia

The description of the vast resources, palaces, and roadways of Atlantis recall Herodotus's description of the assets of the Persian Empire, also borne out by archaeology. The Persian Empire, established by Cyrus the Great in the sixth century BCE, covered over 1 million square miles of territory, including holdings in Africa, Europe, and most of western Asia. The Persian kings developed a vast roadway system to facilitate troop movement and communication between the cities of western Asia. The Behistun inscription lists the immense resources controlled by the Persian king, along with the great number of cultures subject to his rule. Cyrus and Xerxes were both known for their attempts to build canals for military purposes. Herodotus, the father of history and the father of lies, is our main source for the history of the wars between the Persians and the Greeks. He was a fifth-century Athenian whose work was well read.

Atlantis, according to Plato, was a mighty power, vast like the Persian Empire. Atlantis had every resource available: animal, vegetable, and mineral. They even had elephants, an exotic animal to the Greeks, used extensively by the

Persians in war. Just as the kings of Atlantis expected various regions of their empire to support their military, so too did the Persian king, who established troop levies for every satrapy. The well-organized Persian Empire collected taxes in produce, labor, horses, coin, and even mud bricks. The great palaces of Atlantis remind us of the several palaces maintained by the Persian kings throughout Asia at Susa, Babylon, and Persepolis. In Atlantis, the palaces included temples, fountains of hot and cold water, and delightful gardens. Impressively built and richly furnished, the palaces of both Atlantis and Persia defied description.

According to Plato, Atlantis made an "unprovoked" expedition against Europe and Asia 9,000 years before Solon's day. In all too real hostilities, the Persians twice attacked the Greek mainland (in 490–489 and 480–479 BCE) in an act of revenge against Athens's meddling in the Ionian revolt against Persia in 499 BCE. The kings of Atlantis built canals for defense and for communication. Xerxes attempted to make his fleet's crossing to Greece easier by digging a canal across the Chalcidice—the remains of which can still be seen today. Athens headed up the Delian League, a defensive alliance of Greek city–states established in 478 BCE. Athens built long walls to connect her city to her port at Piraeus. This extraordinary defensive building project is echoed in the great defensive wall surrounding the island of Atlantis. The Delian League of Athens aimed to defend against further Persian aggression, free all Greek city–states under Persian rule, and restore all the sacred places destroyed by the Persians.

By 462 BCE, the Delian League successes in the Aegean forced a truce with the Persians. Athens, having come to dominate the league, insisted that the members continue to join forces to preserve the terms of the treaty. Several of the original members of the alliance tried to secede, only to have Athens force their compliance. The Delian League became the Athenian Empire, collecting tribute to support imperial enterprises. Did Athens, like Atlantis, lose her "divine portion" to the more base ambitions of human nature? Or was she, like Plato's victorious Athens against Atlantis, the leader of the Hellenes, who stood alone, "when the rest fell off from her," who "generously liberated all the rest" under the thumb of the ruthless Atlantis (Persia)? Athens's treatment of her league members might be deemed hubristic, even unjust. Was Plato attempting to rewrite history or simply to invent a more just image of Athens as in earlier days?

Sparta

The Athenian Empire was challenged by the Spartans for most of the second half of the fifth century BCE. The Socratics admired Sparta for her political stability. The people of Atlantis were ruled by hereditary kings and princes. Although absolute authority was held by the descendants of Atlas, the elder of the eldest set of twins of Poseidon, there was a sense of duality to the power he held. Sparta was ruled by a dual monarchy of hereditary kings who traced their

lineage to Zeus by way of Heracles. Poseidon gave Atlantis her law code. Spartan lore claims that Lycurgus, a great king, had the Spartan law code approved by the god Apollo before passing it on to the Spartan people.

The Spartan warriors who defended the Spartan way of life serve as the model for old Athens's guardians as portrayed by Plato. Xenophon tells us in *Spartan Constitution* that the Spartans trained and lived in common. They outlawed the hoarding of gold and silver and perceived all property as held in common. The Spartans trained and studied war and the Spartan ideal, separating themselves socially and politically from the craftsmen and herdsmen they relied on for survival. Plato's guardians of the perfect state have a passionate yet philosophical temperament; they are highly educated; they maintain order within the state, while protecting it from outside threats. The guardians have great virtue and no love of gold or silver; they hold all property in common. Furthermore, in a nod to the eugenic practices of the Spartans, Plato described a process in the perfect state by which the "good" people were to be "interbred" with other "good" people to produce the best offspring. The "bad" sort were bred with other "bad" sorts and settled into performing the more subservient tasks of society.

The Spartans were an interesting throwback to the Homeric ideals of the Bronze and Archaic ages. They held on to kingship and a warrior ethic long after both had fallen out of fashion in the democratic, citizen soldier states of the fifth century BCE. Sparta ultimately won the Peloponnesian War against Athens, but the decline of that great state was already evident to Plato and his students.

The Ideal State

As told by Plato, Atlantis was an ideal utopian state, hugely successful until the "divine portion" of her leadership gave way to human baseness. The ideal political organization described by Plato in the *Republic* and the *Timaeus* bears remarkable similarities to Athens's political history, the utopian Atlantis, and the utopian Athens of old. The whole discussion of Atlantis comes after Socrates has regaled his students with a description of the "most perfect" state. According to Socrates, all members of the perfect state are divided into classes and assigned employment based on their nature. Included are the herdsmen, artisans, and guardians. Critias notes how remarkable the similarities are between Solon's Atlantis and the teachings of Socrates on the ideal state, calling them a "mysterious coincidence." Clearly the "mysterious coincidence" that Socrates's ideal state and ancient Athens's ideal political organization are one and the same is a rhetorical device created by Plato to repeat and, thus, reinforce Socrates's ideal. Atlantis's kings are philosopher kings. Furthermore, the story serves to demonstrate that, unlike the Athens of Plato's day—defeated and demoralized—the Athens of 9,000 years ago was a golden age of political organization that Athens of the fourth century BCE should aspire to recover.

The Undersea Kingdom

Over time, the myth of Atlantis has become associated fancifully with stories about an undersea kingdom where humans—or their descendants—continued to thrive, often by breathing water. These stories probably originated with folktales about mermaids and naiads, and by the 19th century, the interest in ancient civilizations inspired many mystics to claim connections to Atlanteans (through channeling, reincarnation, knowledge passed down, etc.) as well as Egyptians and Sumerians. In the early 20th century, pulp magazines popularized the notion of Atlantis as a secret undersea land of water-breathers, and in comic books, Namor the Sub-Mariner, Aquaman, and Superboy's mermaid girlfriend Lori Lemaris all had connections to Atlantis.

Finally, we should note that as we read the *Timaeus* and the *Critias*, there is a sense of competition among Socrates's students to not only demonstrate that they have mastered their teacher's concepts but also that they are good-naturedly competing with one another to provide the most entertaining story. Timaeus's creation epic is monumental in scope, demonstrating that the Greeks are not so immature as Solon's Egyptian priest might think. Critias's story of a fabulous and, ultimately, ideal utopia is filled with drama but also serves as a model of Socrates's teaching from the previous day.

Conclusion

The story of Atlantis has intrigued Western minds for centuries. Atlantis fever spiked in the 19th century with the theories of Ignatius Donnelly, a former U.S. congressman, whose *Atlantis: The Antedeluvian World* (1882) claimed Plato's Atlantis was historical and connected with the Garden of Eden. Donnelly served in Congress just as the United States, a product of global imperialism, began to become an industrialized and imperialistic power herself. Such a young and growing country in an age of progress naturally looks to the past for guidance and to ascertain her place in the grand scheme of history. There are other romantics who claim the Atlanteans sailed to North America and created new civilizations in the Great Lakes region. Others suggest that the sailors of Atlantis transmitted Old World culture to the New World, intimating in the process that those of us with traditions in the Western Hemisphere could not have developed "civilization" on our own. Some even contend that the founders of Atlantis represent a colony from outer space, suggesting in the process that the architectural and military capabilities of Atlantis could not be products of human endeavors. After all, could our forefathers really have been so capable?

Those familiar with Plato and the ancient contribution to global history are at once amused and outraged by such speculations. The same culture that gave

us Euclid and Archimedes also gave us wonderful ideas about human nature that we categorize as political theory and mythmaking. Such far-flung theories rooted in modern prejudices about ancient peoples cannot be supported by a careful reading of Plato or even a layman's appreciation of history and archaeology.

All the actual civilizations noted in this section can be discerned through readings of primary sources created by the people who lived them and a general perusal of the archaeological remains recording their existence. Minoan Crete thrived as a culture, as did Persia and Sparta and Athens. One can argue that the absence of archaeological evidence for Atlantis is no evidence that the civilization did not exist. That is always a possibility. But when it comes to Atlantis there is *no* argument from silence. Plato speaks volumes. Atlantis did not thrive as an actual civilization; Atlantis thrived as the main course of a philosophical "feast," a pretty story to help elucidate Socrates's political theory while offering a brighter picture of the Athens of old, engaging in a selfless, heroic act of war. We as students of the past and living beings of the present take comfort in the hope that such states as Atlantis and the fabled Athens of long ago actually existed; that human beings joined together in peace and justice to create great things in common and to prevent injustices of states that had grown too large and aggressive. Plato paints a beautiful and "perfect" past for Athens at a time when Athens had little to boast of. While the story was a lovely way to teach a lesson and spend an afternoon, we will do well to end our discussion of the historical existence of Atlantis by remembering that in Greek, the word utopia means at once, both "good place" and "no place."

References and Further Reading

Bailey, Harold, ed. *The Cambridge History of Iran*. New York: Cambridge University Press, 1993.

Branigan, Keith. *The Foundations of Palatial Crete: A Survey of Crete in the Early Bronze Age*. London: Routledge & Kegan Paul, 1970.

Burkert, Walter. *Greek Religion*. Cambridge, MA.: Harvard University Press, 1985.

Donnelly, I. *Atlantis: The Antediluvian World*. New York: Harper, 1882.

Doumas, Christo. "The Minoan Eruption of the Santorini Volcano." *Antiquity* 48 (1974):100–115.

Dunstan, William E. *The Ancient Near East*. Fort Worth, TX: Harcourt Brace College Publishers, 1998.

Evans, Arthur J. *The Palace of Minos at Knossos*. London: Oxford University Press, 1925–35; reprint 1964.

Hammond, N. G. L. *A History of Greece to 322 B.C.* Oxford: Oxford University Press, 1987.

Jowett, Benjamin, trans. *The Dialogues of Plato*. Chicago: Bantam, 1952.

Kagan, Donald. *The Fall of the Athenian Empire*. Ithaca, NY: Cornell University Press, 1987.

King, L. W., and R. C. Thompson. *The Sculptures and Inscriptions of Darius the Great on the Rock of Behistûn in Persia*. London: Longman's, 1907.

Lendering, Jona. "Behistun." Available at www.livius.org/bebm/behistun/behistun03.html (accessed May 26, 2010).

Luce, J. V. "Thera and the Devastation of Minoan Crete: A New Interpretation of the Evidence." *American Journal of Archaeology* 80 (1976): 9–16.

Marinatos, S. "The Volcanic Destruction of Minoan Crete." *Antiquity* 13 (1939): 425–39.

Meiggs, Russell. *The Athenian Empire*. Oxford: Clarendon Press 1972.

Pomeroy, Sarah B., et al. *Ancient Greece: A Political Social and Cultural History*. Oxford: Oxford University Press, 1999.

Tavlos, J. *Pictorial Dictionary of Ancient Athens*. London: Hacker Art Books, 1971.

8

Lemuria is not the invention of religious enthusiasts, but rather, actually existed.

PRO   László Kocsis
CON   Claire Brennan

PRO

Another sibling of the Atlantis controversy is the existence or inexistence of Lemuria or of Mu, another lost continent(s) and civilization(s), this time in the Indian Ocean or in the Pacific. The positive theories affirm that these were continents existing in ancient times, which sank beneath the ocean as a result of gradual geologic transformations or because of a sudden, cataclysmic change. Lemuria is the name of a hypothetical lost continent, located in the Indian or Pacific Ocean, submerged around 16,000 BCE, well before Atlantis in 9500 BCE.

The concept originates mainly in the 19th century when serious discontinuities in biogeography were discovered, which could have been only solved by admitting the earlier existence of a continent that split into pieces, most of it sinking beneath the ocean, while the remnants share that common zoological heritage which lies at the origin of its idea of existence. The concept of Lemuria began with Ernst Haeckel, a German zoologist who proposed the existence of Lemuria as an explanation for the absence of "the missing link" fossil records between the primates and humans, which he called the halfway step between apes and humans and named *Pithecanthropus alalus*, translated as "the ape man without speech." Although the missing link, *Homo erectus*, was later discovered by Dubois, first in Java and in East Asia, it created the impression that the first upright human appeared there some 1.8 million years ago according to existing fossils. Later finds occurred in Africa and Europe as well, but Haeckel located the origins of the human species on this lost continent of Lemuria, already prior to that. Lemuria served as the land bridge between the apes of Africa and those of Indonesia and modern man, claiming that the fossil records could not be found because the continent had sunk beneath the sea (Haeckel, 1868).

The continent's name, however, was given later, in 1864 by English zoologist Philip Lutley Sclater, due to the wide spread of its icon species: the lemur, a primate, which would be found from Madagascar to Malaysia, emphasizing the zoological coincidences between Madagascar and India. The name lemur is derived from the Latin word *lemures*, meaning "spirits" or "ghosts of the

179

Lemurs

The name lemur comes from a Roman belief that during the night, your spirit (or lemur) could become restless and this was the cause of nightmares. Crassus and other famous Romans sometimes decided on their actions based on this belief. This is not that dissimilar from beliefs in other cultures, such as the Malay view that a spirit from your body wanders around during the night and your dreams are its experiences.

The name lemur was then applied by Carl Linnaeus to the Red Slender Loris, which he called the *Lemur tardigradus* (it is now known as the *Loris tardigradus*) because most lemurs wandered around at night, and the facial expressions of some of them made them seem as though they were from the spirit world. By tradition, the Aye Aye, the rarest of the lemurs still in Madagascar, has been connected with some superstitions. It has long fingers, and many people on Madagascar felt that if it pointed its finger at a person, that individual was likely to suffer. For this reason the Aye Aye was hunted almost to extinction.

---

night," referring to the monkey's large, reflective eyes and their wailing cries. Though nowadays living lemur species are only found in Madagascar and several surrounding islands, the biogeography of lemurs extends from Madagascar to the Indian subcontinent and onward to Malaysia.

In parallel emerged a circle of adherents of Lemuria, claiming that it was not only a continent but also an ancient civilization, which existed prior to and during the time of Atlantis. Its location alternates; it is believed that it existed either in the Indian Ocean or in the southern Pacific. The location in the Pacific gave birth to the alternative name of the "Motherland" or "Mu," which is thought to have been an enormous continent existing from ca. 98,000 to 8,000 BCE, appearing long before the birth of Atlantis.

The Lemurian Age

The start of the Lemurian Age is placed at the end of the Hybornean Age, which followed the tragic end of the Polarian Age, coinciding with the end of the age of dinosaurs, about 65 million years ago. This didn't mean that saurians could not have and perhaps did indeed linger into the following age, sharing the planet with mankind for quite some time. It is considered that humans appeared much earlier that the Hybornean Age, but became similar to the modern man only by that time.

According to the believers, at some point of evolution, this missing link may have appeared as the Lemurian root race of humans. Due to many climatic and geographic changes and the subsequent changes in fauna and flora, during this Lemurian Age, the early humans evolved into different Lemurian subraces, and the final transition to this root race was marked by the permanent settlement of

humans, thus ending a purely nomadic way of life. Many scholars considered the inhabitants of Lemuria as one of the master races of humanity, as the ancestors of the inhabitants of Atlantis or Atlanteans. The latter were in turn the forefathers of the Aryan race, which emerged from that subrace of Lemurians, which didn't degenerate throughout its historical evolution.

Starting from the 19th-century rise of nationalism in Europe, those ethnologists and linguists who became "racial theorists" argued that humanity is divided into biologically distinctive races, one of them, the most valuable Aryan. The denomination was borrowed from the old Sanskrit language of India, meaning "noble" or "honorable." In particular it was claimed that the Germanic people are the pure and obviously nondegenerated descendants of the proto Indo-Europeans, suggesting that the "Nordic race," characterized by blonde individuals of pale skin, are innately superior to all others. Automatically this line of thought suggested that the "Aryan-Nordic" or "Nordic-Atlantean" race is the master race, the only human race that can and is willing to create civilizations. The other races were and still are considered inferior, simple consumers of these civilizations, which had or will ultimately destroy these cultures. These arguments were originating in the works of the German-Russian-American medium, philosopher, and spiritualist leader of great influence Helena Blavatsky, who postulated that the "Aryans" are the fifth of her seven "root races" (Blavatsky 1888). Her third root race lived in Lemuria, being described by her as hermaphroditic, egg-laying, mentally less developed than the subsequent races, which may signal a reptoid appearance. As a counterbalance she considered that these people were spiritually purer than the following root races. The spiritual but also intelligent Lemurians established trust between them and most animal species, thus domesticating them. As some Lemurians turned to bestiality, as a punishment the gods sank Lemuria into the ocean and created on Atlantis roughly 1 million years ago the "fifth root race" more endowed with intellect. Thus Blavatsky concluded that the present Aryan race still coexists with the fourth root race from a half a million years ago, and that's why some present inhabitants of the Earth are "degenerated spiritually" while others simply became "semianimal creatures."

Although the Lemuria continent argument suffered considerable weakening after the acceptance of actually dominant theories of plate tectonics and continental drift, modern anthropological theories using phylogeography proposed two competing hypotheses about the origins of our own species, the *Homo sapiens*, based on the dispersal of skeletal fossils and ancient human remnants: the *single-origin hypothesis* and a *multiregional hypothesis*. The first hypothesis, also called the *Out-of-Africa with replacement* model, asserts that the last emigration out of Africa occurred around 100,000 years ago and as a result the modern humans displaced all previous human populations in Eurasia, like the *Homo neanderthaliensis*, which were present due to earlier waves of emigration out of Africa. The multiregional theory claims that individuals from the recent

expansion from Africa, all representing *Homo sapiens* or the "man who thinks," intermingled genetically with those human populations of more ancient African emigrations. The recent phylogeographic and genetic findings emphasize Africa as the cradle of humanity, conferring a central role to the single origin hypothesis but also some consideration to the multiregional hypothesis.

Human population might have been reduced to only 1,000 breeding pairs, according to the Toba volcanic catastrophe theory, the eruption in Sumatra some 75,000 years ago. Due to this event, the genetic variety of present-day humans is remarkably small; the closer the group lives to Africa, the higher is its genetic variety. Recent archaeological findings suggest that in fact modern humans survived relatively unscathed in some settlements of India, which support the Lemurian origin hypothesis. The analysis of the genetic diversity in *Helicobacter pylori*, the bacteria causing stomach ulcers that infects 50 percent of the global population, revealed that central place of origin: East Africa, based on the assumption that humans were already infected with these bacteria prior to their widespread emigration some 58,000 years ago.

Historical Accounts of Lemuria

The English theosophical author W. S. Elliott described the continent of Lemuria as stretching from the east coast of Africa across the Indian and the Pacific Oceans. He considered that a set of islands stretching from India to Madagascar—the Seychelles, Mauritius, the Comoro Islands, a line of coral atolls and banks, including the Laccadives, Maldives, the Chagos group, and the Saya de Malha—all indicate the existence of a submerged mountain range or ranges, which possibly formed the backbone of a tract of later Paleozoic-Mesozoic, and early Tertiary land, proposed to be called Indo-Oceania or Lemuria. He claims that Lemuria existed starting from the Triassic and continued throughout the Jurassic and Cretaceous when the third (Lemurian race) coexisted with the famous dinosaurs, living in pine and palm forests. He believes that such humans still populate the Earth today, being largely recruited from the animal kingdom, pointing to the Aborigines of Australia, the Andaman Islanders, some hill tribes of India, the forest people of Ceylon, and the bushmen of Africa. According to the theory, the Lemurian continent was full of active volcanoes, the central mountain range being the most active, as evidence stands the chain of active underwater volcanoes. These volcanic eruptions caused fire, and their subsequent suffocation killed most of the Lemurians. Following this intense volcanic activity, the continent had finally sunken due to tectonic movements. The origins of the human sin leading to the destruction of the continent are also to be found in the act of some of the degenerated Lemurians who created ferocious man-animals, while the intelligent survivors founded Atlantis and subsequently the fifth root race and the Aryan subrace.
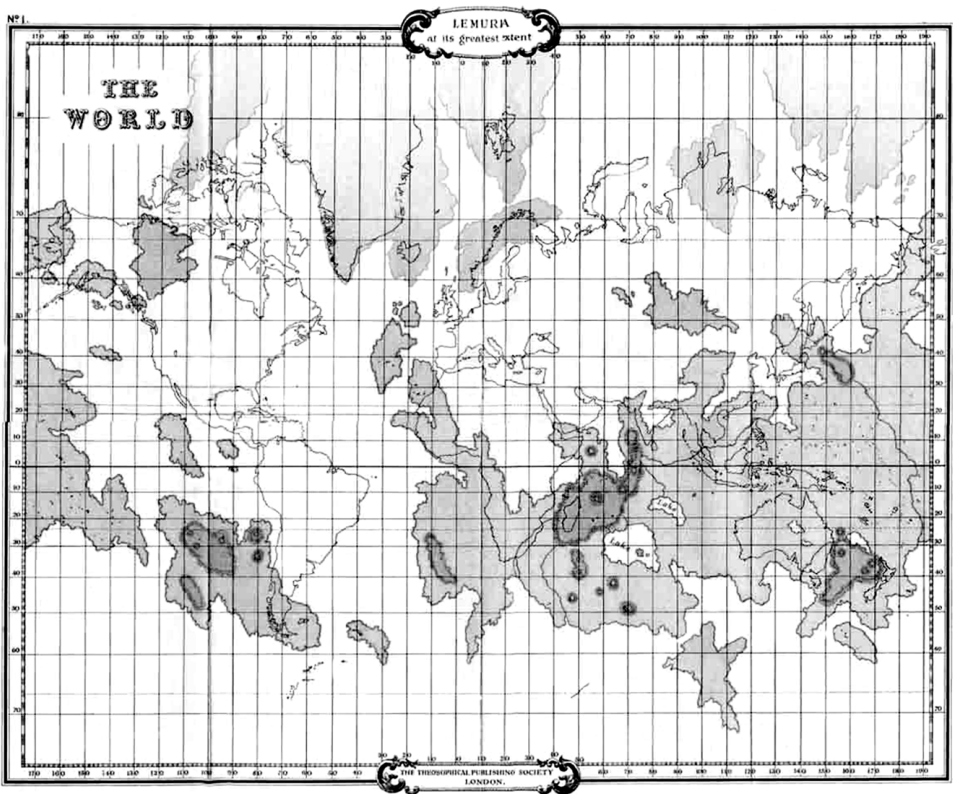
Theosophical map of Lemuria, juxtaposing the present oceans and continents with how the Theosophists envisioned the distribution of land and water over one million years ago, when the Lemurians lived. (W. Scott-Elliot. *The Lost Lemuria*. London: Theosophical Publishing Society, 1904)

In W. Scott-Elliott's (1904) theosophical map, accounting for a continental positioning existing more than 1,000,000 years ago, Lemuria was found in the Indian Ocean, marked with red. James Bramwell (1974) located Lemuria in the south Pacific Ocean. Another British occult writer of the 19th century, engineer J. Churchward (1931), took his inspiration from Indian priests while being a tea planter on the island of Sri Lanka (Ceylon). He identified Lemuria with Mu, in fact another hypothetical lost continent considered to be the original cradle of mankind, stretching from Micronesia to Easter Island and Hawaii, and subsequently the Pacific islanders are supposed to be its last survivors. This location of Mu is, however, contradicted by the oral historical traditions of the Polynesian people and by the absence of large continental masses in the area, since the Pacific basin may represent the missing land mass that gave birth to the moon.

A much better candidate to be the remnant of Lemuria is Sundaland, since it is part of the Indonesian continental shelf, and due to its high elevation it was possibly dry during the Ice Age, forming a land bridge that enabled humans to

reach Australia. As Sundaland slowly submerged, volcanic eruptions like the Toba Lake catastrophe took place there, producing the presumed "population bottleneck," described earlier.

Blavatsky, author of the *The Secret Doctrine*, one of the foundational works of the theosophical movement, claimed to have been inspired from a manuscript of the pre-Atlantean Book of Dzyan found while studying in Tibet. The Book of Dzyan is a reputedly ancient text of Tibetan origin, related to the main Buddhist teaching of the Kalachakra Tantra and also to Chinese taoism. She claimed that it was kept hidden from profane eyes by the initiates of the Great White Brotherhood, but had been shown to her by the Mahatmas (saints in Hindu). The work had originally been written in the sacred language of Senzar, but her critics were skeptical of her having read anything like that, although Max Müller, a great German philologist and orientalist, one of the founders of Indian studies, creator of comparative religious studies, and the translator of the *Sacred Books of the East* from Sanskrit, is reported to have said that she was either a remarkable forger or she had made the most valuable gift to archaeological research of the Orient.

The Vedic language, Sanskrit, was thought to be the oldest of the languages, and scientists like Müller considered that it could help to reconstruct the earliest form of the proto Indo-European root language, enabling us to understand the development path of the classical European cultures, as descendants of the Vedic culture. Blavatsky was charged by some with second-degree plagiarism for quoting primary sources without acknowledging the secondary sources from which they came. However, important ufologists claimed that the Book of Dzyan had originally been produced on the lost continent of Atlantis, which would be another presumed evidence of a previous civilization of Lemuria, which was carried forward by Lemurian descendants: the Atlanteans.

Possible Evidences for Lemuria

Fossil Evidences

Although living lemurs are found mainly in Madagascar and on its surrounding islands, fossils of lemurs abound in the Indian subcontinent and also in Indochina. Since none can be found in the Middle East or Africa, Haeckel proposed the existence of Lemuria, a large continent stretching from Madagascar to India and beyond to Indochina, which was the possible land bridge that enabled the migration of the lemurs (Haeckel, 1868).

Later a synthetic science—biogeography—emerged, which concerned the study of the distribution of biodiversity over space and time, dealing also with evolution, extinction, dispersal, endemic areas, and vicariance. According to the vicariance concept, new lineages came into being when a continuous population (or species) was divided by a new river, mountain range, or rising sea levels, and

as a result two different populations (or species) are created. The patterns of species distribution then can usually be explained through a combination of historical factors such as extinction, speciation, continental drift, glaciation, and river capture, in combination with the area and isolation of available landmasses and food. Thus the biodiversity or richness of species in a certain area could be predicted in terms of habitat area, immigration, and extinction rate. Climate changes such as the glaciation cycles of the past 2.4 million years, have periodically restricted some species into disjunctive refuge areas, which, due to their restricted ranges, may have resulted in population bottlenecks that reduced genetic variation. When a favorable climate change occurred, it allowed for a rapid emigration of specimens out of the refuge habitat and spread into newly available areas.

So regional and local biodiversity are influenced by history, and the makeup of the regional species pool results from the balance of speciation and extinction. The population composition is influenced by the interaction between local extinction and recolonization. A good example of this is the existence of the orangutans in Indochina where they do not show up in India and the similarity of the mammals of the Indian subcontinent with that of Africa. All these show that India and Africa, including Madagascar, were one landmass before or that another landmass connected the eastern shores of Africa with that of India and Indochina.

A good testing ground for biogeography was the Mascarenes Islands in the 16th century until the European settlement, as no peoples were known to exist there before, so much of the islands' original wildlife was still flourishing during the early days. These islands had no native mammals, except for bats. After human settlement, only 16 endemic bird species survived, but 15 became extinct later, since many of them evolved into flightless forms as there were no predators. The most famous extinction was that of the dodo bird, a flightless giant pigeon of the Mauritius. The dodo, standing 1-meter high at 20 kg, fell prey to hunting and the destruction of its habitat as well as plundering of its nests by human-introduced predators like pigs and monkeys. Some other extinct and related species are the Rodrigues solitaire, and the Réunion flightless ibis, constituting the Raphinae subfamily in the pigeon family.

Another great loss encompasses more species of giant tortoises, which comprised the genus *Cylindraspis* and is found today only on the Galapagos Islands. There are 13 living endemic reptile species, including a number of species of day geckoes, but most of the original fauna and flora became extinct since the settlers cleared most of the forests and introduced exotic animal species, such as rats, cats, monkeys, and mongooses.

Genetic research based on their remnants from the 18th century fostered their close relationship and suggested that the Nicobar pigeon, or *Caloenas nicobarica*, from the Nicobar Islands and nearby southeast Asia, which belongs to the Australian, Pacific, and Southeast Asian pigeon family, was the closest living relative of the dodo, but genetic evidence available today only partially

supports this theory. The dodo and solitaire have been linked variously with birds of prey, parrots, and even with the ostrich and emu, but their DNA comparison showed that the dodo evolved away from its relative, the solitaire, due to its geographic isolation about 25 million years ago, long before Mauritius became an island, which is an indirect support for the existence of Lemuria.

Mythical Evidence

In the reptilian literature, the sunken continent of Lemuria or Mu is sometimes shown as the homeland of a reptilian race of creatures, often identified with dragons. Reptilian humanoids, snakepeople, or lizardmen are common motifs in folklore, crypto-zoology, science fiction, mythology, and ufology. They are presented as being evolved on Earth parallel to mankind or as the remains of a pre-human civilization, eventually extraterrestrial. The myths and legends of many cultures throughout history abound in reptilian humanoids; in the Indian Naga, reptilian beings are said to live underground and interact with human beings on the surface. It is said that they once lived on a continent of the Indian Ocean that sank beneath the waves, which points to Lemuria. Indian texts also refer to a reptilian race called the Sarpa. Perhaps the supposedly best known of all "lives" on the other side of the Pacific; it highlights the early prehistoric connections between India, America, and Eurasia. The Aztec main god and sky god, Quetzalcoatl, the "feathered serpent," was variously depicted either as a man, a serpent, or a reptilian humanoid.

Similar impersonation appears in the Bible in the book of Genesis, when God punishes the serpent for deceiving Eve by decreeing, "You should go on your belly from now on," implying that the serpent had legs before then. Thus the snake is therefore often portrayed in Europe as a woman with a snake's tail. East Asians venerate the dragons, and the underwater realms are referred to as where the dragon kings and their descendants live. The main mythology states that there is a direct human lineage descending from dragons, claimed often by East Asian emperors, who were believed to be able to interchange from humans to dragons at will. This lineage is not solely characteristic of Asian mythology; Greek mythology also abounds of reptilians like Cecrops I, the mythical first king of Athens who was a combination of man and snake. Lamia, a child-devouring female, was half woman–half serpent, and the god of the cold north wind, Boreas, was a winged man bearing snakes. Paleontologist Dale A. Russell came up recently with the hypothesis that the Chicxulub meteorite, which left a crater 180 km wide on the Yucatan Peninsula and on the surrounding seabed some 65 million years ago, did not exterminate all the dinosaurs and some of them became intelligent bipods called troodontids, who had fingers and binocular vision similar to humans, enforcing the rationale of the mythical persistence of such creatures (Russell 2009).

Geologic Evidence

India and Meganesia, comprised of Australia, New Guinea, Tasmania, New Zealand, and New Caledonia, are fragments of the once existing supercontinent called Gondwana. Seafloor-spreading separated these land masses from one another, but as the spreading centers became inactive, they fused again into a single plate. So in the region of question there exist two sunken continents: Zealandia at the western edge of the Pacific and the Kerguelen plateau in the Indian Ocean. There is also another significant plateau near the Rodrigues triple point, where the mantle's thickness is double the average (roughly 20 km), although nowadays there is no known geologic formation under the Indian or Pacific Oceans that corresponds to the hypothetical Lemuria. However, drilling carried out in 1999 from a research vessel discovered that the Kerguelen plateau was submerged about 20 million years ago due to rising sea levels. Samples of a 90 million-year-old sediment, collected from the seabed, revealed the existence of pollen and fragments of wood.

The Kerguelen plateau (continent), which presently is submerged 1 to 2 km below sea level, was formed beginning with volcanic eruptions 110 million years ago. Its size might have been even bigger than it is now, since the Broken Ridge underwater volcanic plateau, located west of Australia, was contiguous with it. Geologic evidence presented layers of soil and charcoal, which prove that this was dry land with flora and fauna. Its sedimentary rocks are similar to the Australian and Indian ones, suggesting they were once connected, possibly forming Lemuria.

Zealandia or Tasmantis, with its 3.5 million square km territory being larger than Greenland, is another nearly submerged continent, with New Zealand being its most notable remnant. It broke away from Gondwana, then from Antarctica, and lately from Australia and became almost completely submerged (93 percent) about 23 million years ago.

An interesting huge geologic formation in the Pacific is the approximately 2 million square km volcanic basaltic Ontong Java Plateau near the Solomon Islands, located close to the Antarctic-Pacific ridge by the Louisville hotspot, formed by a mantle plume, which is a lifting of hot rock from the Earth's mantle. This resulted in the 4,300-km long Louisville underwater chain of over 70 seamounts in the southwest Pacific, stretching to the Indo-Australian plate, and specifically to New Zealand, and which may be connected with other ridges reaching the eastern islands of the Pacific.

The Indo-Australian plate may have been connected to the African plate and the Antarctic plate, forming one plate that could host the Kerguelen continent or similar formations in the past. The theory of plate tectonics states that Madagascar and India were parts of the same continent, and if we accept the Lemurian timeline we may conclude that early humanoids may have lived on

this former continent, sharing the same genetic heritage, although the land itself has now drifted apart.

The clusters of islands filling the Indian Ocean between Madagascar and India remain as possible geologic evidences. The Mascarene Islands east of Madagascar have a common volcanic origin and form a distinct ecoregion, comprising the Mauritius, Réunion, Rodrigues islands, the Cargados Carajos shoals, and the banks or former islands of Saya da Malha, Nazareth, and Soudan. The Mascarene plateau, with an area of over 115,000 square km extending from the Seychelles to Réunion, is another evidence for the submerged Lemurian continent, as it has very shallow waters, having depths varying between a mere 8 meters to 150 meters. The plateau presents banks consisted of former coral reefs, some of which might have been islands in the geologic near past, when sea levels were even 130 meters lower than today.

The Saya de Malha bank (in English *mesh skirt*) is a very shallow bank of 40,808 square km, lying southeast of the Seychelles, which reveals that the whole was above water during the Ice Age. The bank is so proper that it was the site of an attempt to create an artificial island by creating seacrete and biorock. The other bank—Nazareth—has an area of about 11,000 square km (according to some sources, this varies between 7,625 and 26,000 square km). Other remnants of Lemuria may be the islands of Seychelles, Réunion, Zanzibar, Mauritius, and Chagos. Réunion, Comoros, and Mayotte are closer to Africa, while Chagos is in the middle of the ocean, halfway from Tanzania to Java.

Seychelles is an archipelago nation of 115 islands in the Indian Ocean, altogether a mere 451 square km, located 1,500 km (930 miles) east of mainland Africa, northeast of the island of Madagascar. The main islands—the inner ones—are located on a shallow bank called Seychelles bank or Seychelles plateau, while the outer islands are situated at 230–1,150 km from the main island Mahe. The inner, central group are composed of 42 granitic islands called the Granitic Seychelles, which form the northernmost part of the Mascarene plateau, all being fragments of the ancient supercontinent Gondwana and thought to have been separated from other continents 75 million years ago. The Outer Islands comprising 46 percent are the Coralline Seychelles, five island groups made up of low-lying coral islands with dry, infertile soils. The fifth, the Amirantes Island, stretches at a distance of about 155 km, all on the shallow Amirantes bank/plateau, with depths varying between mostly 25 to 70 meters.

The Chagos group consists of 60 islands and seven atolls, having an area of around 15,000 square km, of which 12,642 square km form the Great Chagos bank, including lagoons. The Pitt bank, with almost 56 km in length and a width between 20 and 30 km, of an area of 1,317 square km and depth varying between 7 and 44 meters, makes the third largest atoll structure in the world.

The Maldives encompass 1,192 islets and 250 islands, being the lowest lying country with a maximum natural ground level of only 2.3 meters above

sea level. This must have been much bigger in the period of glaciations, so this may be a reference to the classical Sanskrit texts dating back to the Vedic times mentioning the "Hundred Thousand Islands" (*Lakshadweepa*). This generic name, which would include not only the Maldives but also the Laccadives and the Chagos groups, is evidence that it was known and inhabited since ancient times, very possibly before the presumed sinking of Lemuria in 16,000 BCE, and it has been made part of it.

Lakshadweep or Laccadives/Minicoy/Amindivi Islands (the "hundred thousand islands") are located between Arabia and India. It officially consists of about 36 islands and islets covering in total 28 square km, but it also comprises 12 atolls, three reefs, and five submerged banks. Two banks farther north are not considered part of the group—the Angria and the Adas banks—but they have considerable size. Angria bank is a big, shallow, sunken coral atoll on the continental shelf, off the west coast of India, 40 km long by 15 km wide, with a minimum depth of 20.1 meters. Between Angria and the Laccadives lies the Adas bank, with 70 meters at its shallowest point, but all these could have been islands during the last Ice Age, with even more uplift in the previous ages, subsequently of much bigger size when Lemuria presumably existed.

Although current plate distribution may suggest the opposite, the flora and fauna of a portion of land on one plate may be the same as that of the adjacent land belonging to another plate. A good example is the northern boundary of the Indo-Australian plate with the Eurasian plate, which form the Himalaya and Hindu Kush. Its subducting boundary crosses the ocean from Bangladesh to Burma, Sumatra, and Borneo and is not parallel with the so-called Wallace line, which is the biogeographic boundary between the Asian and Australian indigenous faunas. That is also the case for Madagascar and India; although separated now, they preserved their original fauna or part of it in fossils, which is another indirect evidence for the possible existence of Lemuria.

Historical and Legendary Evidence

Lemuria and Kumari Kandam

South India's Tamil Nadu province and the Tamil region of Sri Lanka conserve the myth of Kumari Kandam, the name of a legendary sunken kingdom. One account for this is Blavatsky's description of the dark-skinned Lemurians (the "third race"), showing a similarity between the main inhabitants of southern India, the Negroids, and the Aborigines of Australia, Papua, and Melanesia, all being inhabitants of the former continent of Lemuria. Classic Tamil literature such as the Cilappatikaram places the original ancestral land of the Dravidians south of present day Tamilnadu, in a land that became submerged following a series of floods. The epics also mention the existence of a submerged city called Puhar and Mount Meru. On the other hand, Mount Meru is a sacred mountain in

Hinduism and Buddhism and is considered to be the center of the universe, being located at or in the middle of the earth. The Tamils believe that Mount Meru lies on the sea floor southward from Sri Lanka.

One can find identical geographic denominations in Africa as well: a Mount Meru is located in Tanzania, opposite to the Kilimanjaro, and there exists also a town of Meru located at the foot of Mount Kenya. All this identity of geographic denominations may be more than a simple coincidence—it can serve as a piece of evidence of their relation to the old Sanskrit names and the correspondent Hindu and Buddhist mythology. All together, these represent an indirect evidence supporting the hypothesis that people from East Africa and India were influenced by or shared the same civilization some place in time, the source of which was the continent positioned between them, called Lemuria.

According to the Dravidic mainstream thinkers, mankind also had appeared in this region at ca. 500,000 BCE and evolved into a human race or rather subspecies called *Homo Dravida*, which during the existence of Lemuria, gave birth to a great civilization. This was called Kumari Kandam, and it reached its apex by around 50,000 BCE and existed until 16,000 BCE, when Lemuria presumably submerged. Whatever the origins of the Tamil people may be, there is evidence of traces of the lost civilizations in the region as far as the Easter Islands, which may date back to 20,000 BCE. It is considered by some as the relics of a lost Tamil culture. However, placing the origins of the Easter Islands civilization at a location 20,000 km away from the open sea is not that fortunate. It must not be overlooked that sailors could navigate along the equator by hopping from one small island to the next. They claim that the survivors of these Lemurian civilizations traveled around the region and established their language, which was more original and archaic than the Sanskrit, considered to be a semiartificial composite literary dialect, a kind of the Indian Esperanto. Ancient Dravidic is considered by them as being the prototype of the Indo-European languages, which spread from India into Europe.

Anthropological and Theosophical Evidence

According to theosophists, sex differentiation appeared in humans for the first time during the Lemurian age, when they started to have the modern physical appearance. It is believed that some of them evolved further into the present fifth root race, while others maintained their direct lineage of the semianimal latter-day Lemurians. According to Blavatsky and her followers, these are the Aborigines of Tasmania, Australia, and of various islands in Indochina, even an entirely hairy mountain tribe in China, which are considered inferior subraces, dying out quickly due to "natural selection." I might add that they also die out due to the conflict of civilizations, where the more powerful brings in its diseases and coercive societal structure. Unfortunately, the Tasmanians died out as

many other indigenous people did, and the world became less rich in anthropological diversity than it was before.

The main argument of the theosophical secret doctrine that had a profound impact on many influential writers and politicians of the 20th century is that mankind is "of one blood" but not of the same essence. This view is shared by many if not the majority of people even today. Blavatsky and her followers argued that the great intellectual difference between the Aryans, Turanians, and Semites, on one hand, and the savage aborigines of Melanesia, Ceylon, and Africa, on the other hand, is inexplicable on any other grounds than racial and the latter's belonging to the third race. The mainstream argument in this racial doctrine is that no amount of culture nor centuries of training could raise the intellect or behavioral patterns of these human specimens to a minimum acceptable level, since they miss the "sacred spark" of intelligence, while the fourth (Atlantean) and the present-day fifth root races possess it. According to the theory, a considerable number of racially mixed, so-called Lemuro-Atlantean people were born from various interbreedings, which resulted in degenerated semihuman stocks, such as the Negritos, the Andaman Islanders, and the gypsies, although these argumentations are rejected by the mainstream of science.

The mainstream rather stands on the ground of theoretically and ideologically based equality, expressed in the conviction about the existence of universal human rights and the necessity of abolition of all discrimination and segregation. In reality it can be accepted that there are many "subcultures" and "groups of individuals" who do not attain the acceptable minimum modern standards of intelligence and behavior in spite of having been given vast opportunities for education, although they can be of mixed (sub)racial origin and need not necessarily be of Negroid or mongoloid stock. Albeit the theory and some visible evidence shows that human races or subraces cyclically replace one another, a series of other less favored groups vanishes for good from the entire human family, and we cannot assert that mankind is not essentially "of one and the same essence." In this way the racial theory also permits the real life existence of the giant Cyclopes; being the last subraces of the Lemurians, their single eye equated to the eye of wisdom. As the theory goes deeply into the occult, it lies outside mainstream science, but such assertions that legends preserved historical facts better, since written documents relied on one single nonverifiable source (the author) and could have been forged as well, are gaining more and more popularity among respectable historian circles.

Linguistic and Historical Linguistic Evidence

As a possible evidence, Tamil writer Gnanamutha Devaneyan claimed that Tamil is the root language of Earth and that language originated in Lemuria, which was the cradle of civilizations and the place of origin of the language. He claims that the "primary classicality of Tamil" presents features as phonological

simplicity, morphological purity, agglutinative character, absence of morphological gender, traceability, originality, and natural development (Devaneyan 1967). The Tamilic substratum of the Aryan family of languages is evidenced by the presence of the words *amma* and *appa* in some form in almost all widely spoken languages. However, these words are preserved almost invariably as *anya* and *apa* in Uralian Hungarian, which supports the most widely accepted proposal by the scientific community that an Indo-Uralic language family existed once, encompassing the proto-Indo-European and the Uralic languages. Dravidian linguists and a number of Western scientists support the theory of linguistic connection between the Dravidian languages and the Uralic language family. The latter is represented nowadays mainly by Hungarian, Finnish, and Estonian. Confirming this view, Tamil, as one of the main Dravidic languages spoken mainly in south India and eastern Pakistan, is accepted to be related linguistically with the Uralic and Altaic language families, thus belonging to the larger Scythian language family. Scythian is a somewhat collective denomination of non-Semitic and non-Indo-European peoples of Eastern and Central Europe and Western Asia, characterized to be living in a traditional horsemen society and possessing great artistic skills. Scythia extended from northern India across the Hindu Kush to the Eurasian steppes of what is known today as Kazakhstan, southern Ukraine, and Russia and as far west as Hungary and Romania (Scythia minor), and is described in many ancient Indian sources.

Presumably, Dravidian speakers might have been omnipresent throughout India, before the arrival of Indo-European speakers, which now separate them far from the Ural-Altaic speakers. Legends attesting to Scythic ancestry are found among Hungarians, Scottish, Welsh, and some Slavic people, as well as among Yakuts (Sakha) and Kazakhs, but the Ossethians or Alans of the Caucasus are especially considered genetically and linguistically closest to them.

Going on a parallel route, the new Paleolithic continuity theory asserts that the more ancient roots of civilized man can be traced to the Paleolithic era, tens of millennia earlier than the Copper Age or the Neolithic. The hypothesis suggests that the hypothetical Proto-Indo-European language arrived in Europe together with *Homo sapiens* in the Upper Paleolithic. The theory builds on the lexicon of "natural" languages, which may be "periodized" along the entire course of human evolution, and states that archaeological frontiers coincide with linguistic frontiers. This Indo-European common heritage traceable back to Africa can also be seen as a theoretical evidence of a lost continent linking Africa and Eurasia.

References and Further Reading

Alinea, Mario. "Interdisciplinary and Linguistic Evidence for Paleolithic Continuity of European, Uralic and Altaic Populations in Eurasia." *Quaderni di Semantica* 24, 2 (2003).

Alinei, Mario. "Towards a Generalized Continuity Model for Uralic and Indo-European Languages." In K. Julku, ed., *The Roots of Peoples and Languages of Northern Eurasia*, Vol. 4. Turku: Societas Historiae Fenno-Ugricae, 2002.

Andronov, Mikhail S. "Comparative Studies on the Nature of Dravidian-Uralian Parallels: A Peep into the Prehistory of Language Families." Proceedings of the Second International Conference of Tamil Studies, Madras, 1971.

Avise, J. C. *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard University Press, 2000.

Blavatsky, Helena. *The Secret Doctrine: The Synthesis of Science, Religion, and Philosophy*. Exeter: Theosophical University Press, 1888.

Blavatskaja, Elena, *The Secret Doctrine*, London: Theosophical Publishing Company, 1888.

Bramwell, James. *Lost Atlantis*. Hollywood, CA: Newcastle, 1974.

Burrow, T. "Dravidian Studies IV: The Body in Dravidian and Uralian." *Bulletin of the School of Oriental and African Studies* 11, 2 (1944).

Churchward, James. *The Lost Continent of Mu*. New York: William Edwin Rudge, 1931.

Churchward, James. *The Lost Continent of Mu—Motherland of Man*. New York: William Edwin Rudge, 1926.

Condie, K. C. "The Supercontinent Cycle: Are There Two Patterns of Cyclicity?" *Journal of African Earth Sciences* 35 (2002).

Convention on Biological Diversity. "Third National Report for the Republic of Mauritius." Ministry of Environment and National Development Unit in collaboration with the UNEP/GEF. October 2006.

Cranston, Sylvia. *H.P.B.: The Extraordinary Life and Influence of Helena Petrovna Blavatsky*. New York: Tarcher/Putnam, 1993.

Däniken, Erich von. *Gods from Outer Space*. New York: Bantam, 1972.

Dansereau, Pierre. *Biogeography: An Ecological Perspective*. New York: Ronald Press Company, 1957.

Devaneyan, Gnanamutha. *Language Problem of Tamil Nadu and Its Logical Solution*. Madras: Gnanamutha Devaneyan, 1967.

Dominion Post. "Searching for the Lost Continent of Zealandia." September 29, 2007.

Ellis, Richard. *No Turning Back: The Life and Death of Animal Species*. New York: Harper Perrenial, 2004.

Haeckel, Ernst. The History of Creation, translated by E. Ray Lankester, Kegan Paul, London: Trench & Co., 1868.

Johnson, Kevin P., and Dale H. Clayton. "Nuclear and Mitochondrial Genes Contain Similar Phylogenetic. Signal for Pigeons and Doves (Aves: Columbiformes)." *Molecular Phylogenetics and Evolution* 14, 1 (2000).

MacArthur, Robert, and Edward O. Wilson. *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press, 1967.

Pavanar, Devaneya. *An Epitome of the Lemurian Language and Its Ramifications*. Self-published, 1980.

Pavanar, Devaneya. *The Primary Classical Language of the World*. Madras: Nesamani Publishing House, 1966.

Ramaswamy, Sumathi. *The Lost Land of Lemuria: Fabulous Geographies, Catastrophic Histories*. Berkeley: University of California Press, 2004.

Russell, Dale A. *Islands in the Cosmos: The Evolution of Life on Land*. Bloomington: Indiana University Press, 2009.

Ryan, John, George Dunford, and Simon Sellars. *Micronations: The Lonely Planet Guide to Home-Made Nations*. London: Lonely Planet, 2006.

Scott-Elliot, W. *The Lost Lemuria*. London: Theosophical Publishing Society, 1904.

Shapiro, Beth, Dean Sibthorpe Andrew Rambaut, Jeremy Austin, Graham M. Wragg, Olaf R. P. Bininda-Emonds, Patricia L. M. Lee, and Alan Cooper. "Flight of the Dodo." *Science* 295 (2002).

Sri Krishna, Prem, and Sri Madhava Ashish. *The Measure of All Things: In the Stanzas of Dzyan*. London: Theosophical Publishing House, 1969.

Stow, D. A. V. *Oceans: An Illustrated Reference*. Chicago: University of Chicago Press, 2006.

Tyler, Stephen. "Dravidian and Uralian: The Lexical Evidence." *Language* 44, 4 (1968).

Webb, Edward. "Evidences of the Scythian Affinities of the Dravidian Languages, Condensed and Arranged from Rev. R. Caldwell's Comparative Dravidian Grammar." *Journal of the American Oriental Society* 7 (1860).

CON

Lemuria is the name given to an imaginary continent once thought to have disappeared beneath the waters of the Indian or the Pacific Ocean. When Lemuria was first proposed it was as a pathway that might explain the presence of similar species, both fossil and living, on the island of Madagascar and in Asia. As such it was created by a particular moment in science, as deep time opened up and Darwin's published theory of evolution changed thinking about what the distribution of similar species could mean. These changes in the understanding of nature produced a requirement to explain connections between species that had similar forms even when they were widely distributed. Land bridges and

the lost continent of Lemuria enjoyed a short vogue among scientists, before being replaced by the realization that continents cannot sink. During the second half of the 20th century the far more romantic and astounding reality emerged as tectonic plate theory became widely accepted and unknown sunken continents were replaced by the reality of movement and extraordinary rearrangements among the continents we already know.

By the time that Lemuria's foundations in scientific speculation were destroyed, the continent had somehow gained a life of its own. Lemuria had attracted supporters who believed that a variety of texts, archaeological remains, and direct revelations supported its past existence. Much of this activity was based in the United States, and for those proponents of Lemuria the lost continent had sunken beneath the Pacific rather than the Indian Ocean, as a Pacific location would allow a direct American connection. For some of these supporters of Lemuria the sunken continent's usefulness lay not so much in providing a land bridge across the Pacific as in adding weight to the phenomenon of the more famous missing continent Atlantis. At the same time that Americans embraced Lemuria, the continent also gained existence as a feature of Tamil nationalism in southern India, where it still exists within the national curriculum and in official government pronouncements.

The Scientific Origins of Lemuria

During the late 18th and throughout the 19th centuries the sciences of geology and biology were undergoing upheaval. The first major step in that process was the discovery of deep time in the late 18th century when it was realized that the formation of some types of rock, and the manner in which they were arranged, would occur over a large amount of time. Previously the form of the modern world had been explained by a series of catastrophes as Western thinkers considered Earth to be only 6,000 years old (rather than the present accepted age of 4.5 billion years). This age had been calculated using the Bible, estimating the age of Earth from the genealogies contained in the Old Testament and accepting Adam and Eve as their starting point. With a 6,000-year-old Earth, catastrophes were used as a possible explanation for observations such as fossils of extinct species. As the estimation of the age of the Earth increased, other explanations became possible, and in 1832 William Whewell described the slow processes that, over a very long time, could create the physical features of Earth—a process he named uniformatarianism. Charles Lyell made geology into a respectable science with the publication of his three volumes of *Principles of Geology: An Attempt to Explain the Former Changes of the Earth's Surface* between 1830 and 1833. In the process, he publicized the idea of uniformitarianism and introduced the idea of gradualism—that large geologic features could be explained by slow processes operating over a long period of time rather than by catastrophes. Lyell's work influenced that of Charles Darwin and the publication of *On the Origins of Species* in

Philip Lutley Sclater and Lemuria

Philip Lutley Sclater, the zoologist who developed the concept of Lemuria, was the son of a landowner and born on November 4, 1829, at Tangier Park, Hampshire, in the south of England. Educated at Winchester College, he won a scholarship to Christ Church, University of Oxford, where he studied mathematics, natural history, and modern languages. He then started studying law, but it did not interest him, and he began to focus on ornithology.

His interest in fauna took him to the United States in 1856, where he traveled down the Mississippi, and then returned to England where, two years later, he developed his idea of the six zoogeographic regions of the world—linking Madagascar with India because of similarities in the wildlife. He later went back to North America, and also to South America and South Africa.

Sclater's studies were interrupted by having to practice law for a while and also being the private secretary to his brother who had been appointed president of the local government board. Always keen on meeting and helping other zoologists, some found him arrogant and aloof. He died on June 27, 1913, from a carriage accident.

1859, which provided a biological counterpart to the geologic theory of uniformatarianism. One consequence of Darwin's theory of evolution was the necessity of finding physical connections between similar species separated by large distances as such species had been linked until recently in their evolution.

Lemuria emerged from this period of scientific reassessment of Earth, and its name was first proposed by the biogeographer Philip Lutley Sclater in a scientific article published in 1864. In the article, Sclater reviewed the animal species that occurred on the island of Madagascar, which lies about 200 miles from the east coast of Africa. He noted that much of the fauna of the island was very distinct, but that its few similarities were to animals found in South America and India rather than to those on the much nearer African continent. He noted that a number of distinctive and widespread African species (such as antelopes and the hippopotamus) were absent from the island, and that the famous lemurs of Madagascar were quite different from related species found in either Africa or Asia. As a result he concluded that Madagascar had never been part of Africa, but that Africa contained land that had once formed part of Madagascar; that Madagascar and its nearby islands had long been separated from any other landmass, but that some land-connection must once have existed between India and Madagascar; and that there must once also have been a land connection to the Americas. According to Sclater, portions of that lost continent joined with Africa, some perhaps with Asia, and others formed Madagascar and its associated islands. As lemurs are the most famous species endemic to Madagascar, Sclater proposed, in the last sentence of his article, that the missing continent be called "Lemuria."

Sclater was correct in his assessment of the distinctiveness of the animals of Madagascar and in the connections that had once linked Madagascar with Asia and South America, but he was wrong only in requiring a lost continent to explain that position. However, Sclater was not the only 19th-century scientist to consider Lemuria as a possible explanation for observed evidence, and the continent received acceptance in mainstream sources for a while. For example, the German evolutionist Ernst Haeckel proposed Lemuria as the probable cradle of mankind in an 1868 work published in German (and republished in English in 1876 as *The History of Creation*). The idea was also briefly accepted in broader intellectual circles, appearing in Frederick Engels's essay *The Part Played by Labour in the Transition from Ape to Man* and in H. G. Wells's *Outline of History.*

However, even in the 19th century, Lemuria was not always accepted as fact. While the missing continent was at first accepted by Alfred Russel Wallace (the codiscoverer of the theory of evolution), it was later rejected by him because it could not explain the distribution of species between islands in Asia. Wallace rejected the idea of a sunken continent because while it might explain similar creatures now found at a distance from one another, it could not explain the mystery of very different clusters of creatures located very close to one another. Wallace observed this phenomenon in Southeast Asia, for example, in the very different biotas of the islands of Lombok and Bali. While located just over 20 miles from each other, these islands are different in soil type and in the animals endemic to them. The channel separating the two islands may be narrow, but it contains strong currents and is too deep to have been left dry by falling sea levels during an Ice Age. As a result of observations across the region, Wallace proposed the "Wallace line," now recognized as Wallacea—a zone that marks the edge of the Asian continent and its biota. Such an observation cannot be explained by a sunken continent and helped Wallace decide against the existence of Lemuria as an explanation for the distribution of species.

Replacement of Sunken Continents with Tectonic Plates

Lemuria was not a plausible explanation for the distribution of species or of the arrangement of the world, and another theory gradually rose to force its rejection by scientists. A related problem to that of species distribution was the arrangement of the Earth's landmasses and whether there might be continents missing. As early as the 16th century, cartographers had noted that Earth's landmasses appear to fit together along their edges. Nineteenth-century observations of the wide distribution of some kinds of fossils and of some sequences of rock formations again pointed to previous connections between now separated landmasses. Associated with this question of continents was one of mountains and how they came into existence. The most popular 19th-century explanation for mountains was that Earth had been warmer in the past, and that as it cooled it had

contracted, scrunching up its surface into the formations known in the present. This theory allowed for the instability and collapse of continents to form the sea-floor and the raising of areas that had been seafloor to dry land. It also explained the presence of marine fossils on land and of similar fossils wide apart as this process of scrunching deformed Earth's surface unpredictably. And it allowed for the presence of sunken continents.

By the 20th century the theory of cooling and scrunching was challenged by three sets of evidence. The first was the careful mapping of mountain ranges, especially the Swiss Alps and the Appalachians. That careful mapping indicated that there was simply too much rock contained in these mountain ranges to be explained by contraction and scrunching. The second line of evidence came from the Great Trigonometrical Survey of India that took place in the 1850s. During that survey there were problems with measuring the relative locations of places separated by the Himalayas. This led to a separate survey of the gravity associated with the Himalayas and the discovery that in this case there was less gravitation deflection than there should be, indicating that there was less mountain than might be expected. This observation could be explained in one of two ways—either the continents were less dense than the layers of rock below, or the Himalayas had deep roots but floated in the layer below like giant icebergs in water. In either case the gravitational measurements provided clear evidence that the continents were lighter than lower levels of the Earth's crust, and that ocean beds and land masses could not swap places—a theory, now known as isostasy. The third contradictory piece of evidence was physicists discovery of radiogenic heat. The discovery of heating due to radioactive elements simply contradicted the idea of a steadily cooling Earth and so destroyed the roots of the theory of contraction.

As these factors removed the credibility of the contraction and scrunching theory of the Earth's surface, other explanations were sought for the arrangement of the continents and the distribution of species. The first scientist to propose drifting continents as a solution was a meteorologist rather than a geologist. Alfred Wegener, in a series of papers published between 1912 and 1929, proposed that the continents floated around Earth, rather than rising and sinking. This would explain their matching coastlines, allow them to have had different connections in the past, and explain fluctuations in global climate. His theory was widely rejected at the time, but supporting evidence was added by European scientists who suggested possible mechanisms for movements of the Earth's crust. While the theory gained some acceptance in Europe during the early 20th century, it continued to be rejected in North America where land bridges were used to account for perplexing fossil and species relationships instead. The only data available to support the theory of land bridges were the very data that the theory sought to explain, meaning that the argument was circular. However, in North America rising and sinking land bridges were seen as less far-fetched than the drifting continents accepted in Europe.

Yet further evidence for drifting continents continued to emerge. Investigations of Earth's magnetic field in the middle of the 20th century led to interest in its past. It had been noted that as rocks cool and become solid, they reflect the orientation of Earth's magnetic field. Once the rock is solid, that original alignment remains even when the rock is shifted within the magnetic field. This observation led to an examination of the magnetic field orientations of different ages of rock across Earth's continents. Massive amounts of data yielded evidence that the polarity of Earth's magnetic field periodically reverses. That data also indicated that either the magnetic poles wandered or the continents had moved. By comparing the magnetic records of different continents with one another, it became clear that either different magnetic poles affected different continents (a physical impossibility) or the continents had changed their relative positions in the past. Continental drift had been proved in this fashion by 1956, yet North American scientists remained hostile to the idea.

However, further evidence in support of drifting continents continued to emerge. World War II had led to an interest in mapping the ocean floor, supported by government money on both sides of the Atlantic. As a result, evidence emerged that the ocean floor remained geologically active—it was not an inactive sunken remnant as required by theories of sunken continents and land bridges. Work in mapping the oceans finally demonstrated that new ocean floor was formed by the eruption of submarine volcanoes, and that the oceans were being split apart at their centers and sinking back down when colliding with the continents, causing compression. In the Pacific this results in the Pacific Ring of Fire, the circle of volcanic and earthquake zones that ring the edge of the Pacific plate. By the end of 1968, the theory was fully developed and was named plate tectonics. The existence of different arrangements of Earth's continents in the past became evident. The continents were once clustered together in supercontinents, on which related species of plants and animals evolved. As the supercontinents fractured and drifted, they carried these species with them, leading to the presence of lemurs in Madagascar and doing away with the need for Lemuria. The strange truth that Earth's crust is formed of fragments 45 to 60 miles thick that move against one another at a rate of one to four inches a year had finally been accepted in North America. This understanding of the structure of Earth provided explanations for a wide range of phenomena, including volcanoes, earthquakes, and mountains. It also explained the distribution of species, both fossil and living, across Earth—including Sclater's lemurs. Tectonic plate theory does not leave any room for sunken continents.

Lemuria's Connection with Atlantis

During the time between Lemuria's proposal and dismissal in scientific circles, this phantom continent had drifted into stranger company and become established

among nonscientists. In the process, the legend of Lemuria became intimately intertwined with that of another lost continent, Atlantis. While, to paraphrase Lady Bracknell in *The Importance of Being Earnest*, "To lose one continent might be regarded as a misfortune, to lose both looks like carelessness," it did not appear in this light to those linking Lemuria with Atlantis. To them, the loss of a second continent might indicate the feasibility of sinking a large mass of land, and that Lemuria was first proposed by a scientist seemed to add to its validity. Atlantis's genesis was more problematic than that of Lemuria as it was drawn from fragments written by Plato and nothing else. No further evidence supporting the existence of Atlantis has ever been discovered, making the seemingly scientific proof of another lost continent useful. Despite this enthusiasm for scientific support, the evidence advanced by Lemuria's later enthusiasts tended to be of a type unacceptable in scientific circles.

The first major nonscientific appearance of Lemuria was in the writings of Helena P. Blavatsky, founder of a group known as the theosophists in the late 19th century. Blavatsky was originally Russian, but was based in New York by the time she published her writings on Lemuria. Her book, *The Secret Doctrine*, first published in 1888, was based on a set of writings she claimed came from an Indian guru, and on direct revelation. The existence of the guru cannot be independently verified, and handwriting analysis has indicated that Blavatsky herself wrote the work she attributed to him (Nield 2007: 50). Her clairvoyant revelations of Lemuria clearly linked it to the other lost continent, Atlantis. Her writings implicated both lost continents in a scheme of human evolution that involved a series of root races. The third of these root races lived on Lemuria, laid eggs, were hermaphrodites, and were destroyed when their discovery of sex led to bestiality and the displeasure of the gods. The few survivors then moved to Atlantis where they became the fourth root race, had access to airplanes and explosives, and a type of wheat donated to them by aliens. Once Atlantis was destroyed the fifth root race developed, to which modern humans belong. Such a sequence of human evolution is not supported by any physical evidence. Blavatsky also had her Lemurians coexisting with the dinosaurs, although mainstream science indicates that humans did not evolve until long after the dinosaurs were extinct. However, while Blavatsky's writings might seem entirely unbelievable, she attracted a following of 100,000 people and contributed to the continued existence of Lemuria by moving it from the Indian Ocean to the Pacific. Such a location allowed it a connection with North America, and the attraction of considering that North America was integral to the development of humanity and of civilization ensured that the legend lived on.

Blavatsky was not the last of Lemuria's enthusiasts. The great proponent of physical evidence for Lemuria in the 20th century was Lewis Spence. Spence had earlier championed the existence of Atlantis and the association of Atlantis with the somehow weightier continent of Lemuria is promoted in his 1932 book

The Problem of Lemuria. Much of Spence's evidence was drawn from interpretations of archaeological evidence (Spence dismissed the possibility that Polynesian people might be responsible for the stone structures he observed in Polynesia) and on interpreting traditional stories as referring to a vanished continent. However, his work depended on the idea that it has been quite common for Pacific islands to appear and disappear. He interpreted early European sightings of land that could not be refound—such as the 1688 sighting of land by the Dutchman John Davis—not as evidence of the difficulty of seeing islands from a distance, ineffective navigational instruments, and the problem of determining longitude before the availability of reliable chronometers on ships, but as evidence of islands that had actually disappeared. In addition, he claimed that well-known islands had at times been submerged and later reappeared. Modern understandings of the formation of Pacific islands as the result of volcanic eruptions (either directly, as with the Hawaiian chain, or indirectly, as with coral atolls perched on the top of submerged volcanoes) do not allow for them suddenly disappearing. Instead, sightings of phantom islands must be understood as a result of the difficulties of navigation during early European exploration of the Pacific and as the tainting of observation by expectation. Mariners hoped to find islands and sometimes saw them in the distance where they did not exist.

Blavatsky and Spence might be excused their fascination with sunken continents as they wrote before tectonic plate theory was fully developed. They both benefited from publicity in North America, where the idea of continental drift was not accepted and the idea of a local lost continent, and so an association between the Americas and early human civilization, might be attractive. No such excuse exists for David Childress who promoted Lemuria in 1988 in his book *Lost Cities of Ancient Lemuria and the Pacific*. The book is a travelogue of Childress's time in the Pacific, and he finds hints of a lost continent in his observations of the Pacific islands. The book was not presented as a careful academic argument, and it was not one, but it served to defend the notion of a lost continent in the Pacific. Childress's argument relied on an acceptance of changes in scientific understanding indicating that science is unreliable and on the possibility of ancient texts surviving in secret and revealing an entirely unknown continent (1988: 6, 8). The book served to continue the myth of Lemuria, but it presented no verifiable evidence to support its assertions. Instead it promoted Lemuria by means of a charismatic host, noting seemingly ancient artifacts and yarning about the secret history of a lost continent.

Another Lost Continent—Mu

The connection between Lemuria and the later lost continent of Mu is problematic. The names appear similar, the location is again the Pacific, and the style of evidence proposed by the continent's supporters again includes mysterious

This illustration depicts the destruction of the Lemurian continent of Mu by flood, which is to have occurred in approximately 10000 BCE. (Mary Evans Picture Library/The Image Works)

ancient texts, lost languages, and connections between India and the Americas. The vision of Mu put forward by its supporters is similar to that of Lemuria put forward by Blavatsky and Spence. According to its greatest publicist James Churchward, Mu was a vast and pleasant continent in what is now the mid-Pacific. It enjoyed a high level of civilization, peace, and a large population. Megaliths, such as those that remain on Easter Island, were flown into place as gravity had been overcome, and communication over distance was accomplished by inner vibrations. Migrants from Mu were responsible for the civilizations that later developed elsewhere—including those of Atlantis, the Mediterranean, Asia, and the Americas. Lemuria and Mu are not generally identified as the same lost place, but their similarities make it worth examining Mu's shaky foundations in order to finish destroying those of Lemuria.

In theory, the existence of the lost continent of Mu was discovered in Mayan texts read by Charles Etienne. The problem is that Etienne believed he was decoding the tablets, but no key to Mayan writing exists. As a result he applied a letter-based understanding to a system of writing that is not letter based, and then constructed a narrative from the fragments of writing he had available to him. His ideas were picked up in Philadelphia by Ignatius Donnelly, and from him by Churchward, who was Mu's strongest promoter. Churchward's evidence for this continent is not convincing. For example, he "discovered" his continent by deciphering 5,000-year-old clay tablets he claimed to have seen in India, but which have not been verified by any other source. The tablets were written in a language understood by only three people

and, as Churchward explained, understanding the tablets was not simply a matter of reading them. Instead the real meaning of the tablets required interpretation, as "many of the apparently simple inscriptions had hidden meanings" (Churchward 1931: 18). The degree of interpretation necessary is evident from Churchill's explanation of select symbols—a rectangular block was the symbol of the lost continent, and could be interpreted as "Moo, Ma, mother, land, field, country" or "mouth." Churchward's interpretation of images that he claimed to depict the destruction of Mu shows a similar interpretative license. He claimed to have found two depictions of the destruction of Mu, one in Egypt and one in North America. He noted: "There is, however a marked difference in the two. The Egyptian depicts Mu falling into an abyss of fire, the North American Indian tells of the waters flowing in over her and submerging her." Churchward did not see these two completely different depictions as problematic, instead asserting that the images show "two phases of the general description of her destruction—so that both the Egyptian and the North American Indian are shown to be correct. Although these two peoples are so widely separated on the earth's surface now, at that time both knew the scientific cause of Mu's destruction" (Churchward 1931: 67–68). The North American image is depicted in Churchward's book and does not obviously show a land being flooded.

Both Mu and Lemuria owe some of their attractiveness to finding the origin of civilization in the Pacific rather than the Old World. However, the way in which evidence is drawn from remaining artifacts in the Pacific is problematic. Churchward wrote:

> Throughout the length and breadth of the Pacific Ocean are scattered groups of small islands. On scores of them are the remains of a great civilization. There are great stone temples, cyclopean stone walls, stone-lined canals, stone-paved roads and immense monoliths and statuary—works that required continental resources and workmen of skill. Yet we now find them not on a great continent, but on mere specks of land inhabited by savages and semi-savages. (Churchward 1931: 87)

This passage is an example of the unthinking racism required to depict the societies of Polynesia as degenerate remnants of a lost civilization rather than energetic and adaptable cultures. Pacific islanders and their ancestors were perfectly capable of erecting stone works such as those that have regularly perplexed visitors to Easter Island. Further study of Easter Island, a key example for Churchward, has revealed both the methodology of the erection of the statues and the ways in which the island environment changed, robbing islanders of the resources to maintain their previous lifestyle. No sunken continent was required.

Mu is suspiciously close in form to Lemuria and appears to be a revival of old ideas about sunken continents, which seeks to avoid an association with lemurs and with ideas about the evolution of humans from primates. It does not

seek to use Lemuria's scientific beginnings to claim ongoing scientific credentials, but instead falls within a mystical understanding of prehistory and a similar rejection of mainstream science. It is not supported by tangible evidence of any kind, but by imaginative interpretations of scattered and fragmentary artifacts. It is undermined by the sheer depth of the Pacific Ocean, which does not harbor any lost continent of the magnitude proposed for Mu, or indeed even for the slightly less ambitious Lemuria.

Tamil Adoption of Lemuria

Sumathi Ramaswamy (1999, 2000, 2004) has explored the last existence of Lemuria—the continent lives on among the Tamil people of southern India. As a result of British colonialism, Lemuria made its way to India during its period of acceptance in scientific circles. The existence of a lost continent seemed to constitute a neat explanation of the language division between northern and southern India, with the speakers of Tamil seen as refugees from the large, lost southern continent of Lemuria. It was as a result of British approval that Lemuria first appeared in an Indian textbook, in 1876. Lemuria was swiftly adopted by the residents of southern India, finding favor with Tamil intellectuals from the 1890s onward. A lost continent seemed to match Tamil stories of a homeland that was lost twice over and meshed with Tamil understandings of themselves as different from the people of northern India. As with the promotion of Lemuria in North America, the lure of claiming a close connection with the birthplace of humanity and the cradle of civilization was attractive. As a result Lemuria continued to appear in school textbooks in the 20th century and to become more and more acceptable in public discourse in southern India. As Tamil nationalism emerged during the 1920s and stood in opposition to both British rule and an Indian nationalism dominated by language groups from northern India, the appeal of a Tamil-associated birthplace of civilization only became stronger. After 1950, Lemuria appeared in public speeches and it was the subject of a government-approved documentary in 1981. Some aspects of Lemuria's past, particularly its associations with Sclater's primates and with the strange beings described by Blavatsky, lessened its attractiveness as a homeland, but these were outweighed by the way in which science seemed to lend authority to Tamil traditions. These two competing factors led to a compromise—Lemuria still exists in official textbooks in southern India, but from 1903 it has often appeared under a traditional Tamil name—Kumari Kandam.

Conclusion

Lemuria never physically existed. There is no need for it, no room for it, and no evidence for it. When developments in geology led to theories of drifting

continents (a notion surely more romantic, and yet true), sunken pieces of land were no longer needed to explain scientific observations of the distribution of species. Such observations could instead be explained by the breakup of super-continents and the drift of fragments into their present positions. However, the brief and limited approval given by science to Lemuria lent it weight in stranger circles, even though those circles often explicitly rejected the virtues of science as a way to understand the world. The scientific genesis of Lemuria allowed it a reality denied to its cousin Atlantis, and so supporting the existence and loss of Lemuria supported claims about the reality of Atlantis, and about the significance of the Pacific, the Americas, and India in the history of civilization. These portions of the world have been denied prominence in scientific stories of human evolution and the birth of civilization, and by lending them significance, the lost continent of Lemuria has become attractive in those regions. This attraction has led to a variety of strange bedfellows promoting Sclater's continent, a promotion that continues in official circles in southern India and among those that believe in mysterious texts and revealed knowledge elsewhere.

References and Further Reading

Cervé, W. S. *The Lost Continent of the Pacific*, 6th ed. San Jose, CA: Supreme Grand Lodge of Amore, 1954.

Childress, David Hatcher. *Lost Cities of Ancient Lemuria & the Pacific*. Stelle, IL: Adventures Unlimited Press, 1988.

Churchward, James. *The Lost Continent of Mu*. New York: Ives Washburn, 1931.

Diamond, Jarad. *Collapse: How Societies Choose to Fail or Succeed*. New York: Viking, 2005.

Engels, Frederich. *The Part Played by Labour in the Transition from Ape to Man*. Moscow: Foreign Languages Publishing House, 1949 (originally published in 1876).

Feder, Kenneth L. *Frauds, Myths, and Mysteries: Science and Pseudoscience in Archaeology*, 3rd ed. Mountain View, CA: Mayfield, 1999.

Haeckel, Ernst. *The History of Creation*, translated by E. Ray Lankester, Kegan Paul, London: Trench & Co., 1868.

Howe, K. R. *The Quest for Origins: Who First Discovered and Settled New Zealand and the Pacific Islands?* Auckland: Penguin, 2003.

Lyell, Charles. *Principles of Geology: An Attempt to Explain the Former Changes of the Earth's Surface*. London: John Murray, 1835.

Milne, Anthony. *Doomsday: The Science of Catastrophic Events*. Westport, CT: Praeger, 2000.

Nield, Ted. *Supercontinent: Ten Billion Years in the Life of Our Planet*. Cambridge, MA: Harvard University Press, 2007.

Oldroyd, David R. *Thinking about the Earth: A History of Ideas in Geology*. London: Athlone, 1996.

Oreskes, Naomi. "From Continental Drift to Plate Tectonics." In Naomi Oreskes, ed., *Plate Tectonics: An Insiders' History of the Modern Theory of the Earth* (pp. 3–27). Boulder: Westview, 2001.

Ramaswamy, Sumathi. "Catastrophic Cartographies: Mapping the Lost Continent of Lemuria." *Representations* 67 (Summer 1999): 92–129.

Ramaswamy, Sumathi. "History at Land's End: Lemuria in Tamil Spatial Fables. *Journal of Asian Studies* 59, 3 (August 2000): 575–602.

Ramaswamy, Sumathi. *The Lost Land of Lemuria: Fabulous Geographies, Catastrophic Histories*. Berkeley: University of California Press, 2004.

Sclater, P. L. "The Mammals of Madagascar." *Quarterly Journal of Science* (April 1864): 212–19.

Spence, Lewis. *The Problem of Lemuria: The Sunken Continent of the Pacific*. London: Rider, 1932.

Wells, H. G. *The Outline of History, Being a Plain History of Life and Mankind*. London: G. Newnes, 1919.

9

Native American peoples came to North and South America by boat as well as by land bridge.

PRO   Peter N. Jones
CON   James Seelye

PRO

Ever since Euro-Americans first arrived in the Americas and discovered that they had been inhabited by Native Americans, various arguments have been put forth concerning how, and by what means, Native American peoples came to North and South America. Early theories concerning the peopling of the Americas were largely grounded in myth or legend. In fact, it was not until Thomas Jefferson demonstrated in 1784 that contemporary Native American tribes were responsible for the large mounds found throughout the North American southeast, as opposed to being related to the mythical Atlantians or one of the lost tribes of Israel, that theories became grounded in empirical evidence. However, this did not stop people from continuing to speculate about possible Old World cultures peopling the Americas. For example, during the 19th and early 20th centuries, a wide range of theories were put forth concerning the peopling of the Americas, ranging from the argument that voyagers in Chinese junks crossed the Pacific and reached the Northwest Coast, or that Phoenicians from the Mediterranean made their way across the Atlantic, or that West Africans reached eastern Mexico at some point during the late Holocene epoch.

By the beginning of the 20th century, however, with the development of certain fields of inquiry and the adoption of empirical scientific methods, scholars began to conclude that Native American peoples must have come to the Americas from Asia. Historically, the most commonly accepted means by which this colonization was thought to have taken place was via a land bridge that was hypothesized to have formed intermittently during the Pleistocene epoch (126,000–11,000 years ago) across the Bering Sea, connecting northeast Asia with northwest Alaska. By 1933 enough data had been compiled for W. A. Johnston (1933) to use estimates in his "Quaternary Geology of North America in Relation to the Migration of Man" of a lower sea level to demonstrate that a land bridge had in fact existed between Asia and North America during the last glaciation.

As the melting continental glaciers discharged their stored water into the oceans at the end of the Pleistocene, the sea level rose, flooding the Bering land

207

This fresco at the National Anthropological Museum of Mexico depicts hunters migrating from Asia across at the Bering land bridge. North, Central, and South America were populated over a period of thousands of years after nomadic hunters crossed over a land bridge into North America ca. 35000–9000 BCE. (The Art Archive/National Anthropological Museum Mexico/Dagli Orti)

bridge and establishing the Bering Strait that separates North America and Asia today. This theory, which came to be known as the Bering land bridge theory of migration, was largely based on the faunal data evidence that such migrations took place. For example, by 1938 scholars had generally concluded that, as written by Ernst Antevs in "Review: The Holarctic Region":

> "[T]he subsequent interglacial and glacial ages induced repeated north-and-south migrations of the biota between the refuges and the glaciated areas. There were also west-and-east movements, which resulted in mixtures of the forest biota—but not of the steppe biota—of Eurasia and North America. There was especially a spread from Asia via a Bering land bridge or isthmus to North America." (1938)

It was not a far stretch for scholars, therefore, to conclude that along with various floral and faunal biota, humans also migrated into the Americas via this land bridge. When exactly did the land bridge cease to be a corridor for migrations remained undetermined. By 1967 enough geologic evidence had accumulated, however, to enable David M. Hopkins, in *The Bering Land Bridge*, to speculate that the Bering land bridge was inundated by the rising sea level for the last time sometime shortly after 10,000 years ago.

Clovis-First Theory Debated

Known by various other names, such as the Clovis-First theory or the single-migration wave theory, the general theoretical model that Native American peoples migrated from Asia to North, and subsequently South, America by means of the Bering land bridge at the end of the Pleistocene held sway until multilateral evidence could no longer support it. Historically scholars, as discussed by David M. Robinson and Elizabeth Pierce Blegen (1933) in "Archaeological News and Discussions," thought Native Americans first arrived in the Americas no earlier than 11,500 years ago, based on the distinctive style of fluted lithic point found in 1932 near Clovis, New Mexico. According to the Bering land bridge theory, a small group of Native American hunters, possibly numbering under 100 individuals, entered North America from northeast Asia by means of the Bering land bridge at the end of the last Ice Age approximately 11,500 to 12,000 years ago. Before that, it was believed that most of Canada was covered by the Cordilleran ice sheet centered over the Rocky Mountains and the Laurentide ice sheet centered over Hudson Bay, effectively blocking any way for Native Americans to enter North and subsequently South America. Furthermore, it was argued that glacial ice from various Alaskan and coastal glaciers blocked any passage along the Northwest Coast of North America and that the land bridge was the only viable option. At the end of the last Ice Age, Bering land bridge theorists, according to Robson Bonnishsen and Karen Turnmire (1999), in "An Introduction to the Peopling of the Americas," hypothesized that Native Americans funneled through an "Ice-Free Corridor" as a result of the melting of the two ice sheets into the plains of North America, subsequently populating Central and South America.

Although a widely held theory, the Being land bridge model was not universally accepted, and scholarly debates about the peopling of the Americas continued to be wide ranging. For example, Alex Krieger (1962), in "Comments on 'The Paleo-Indian Tradition in Eastern North America' by R. J. Mason," argued that it was impossible to believe that the Americas were unoccupied before the appearance of fluted points, primarily because there were several known sites that seemed to predate Clovis-era points. These early sites, because they could neither be conclusively dated to the pre-Clovis era nor demonstrated to be of human manufacture, were often ignored by those who argued for the Bering land bridge model of entry. The rigorous criteria for proof demanded to properly evaluate claims of pre-11,500 before the present occupation of the Americas were expressed by C. Vance Haynes in "The Earliest Americans":

> For establishing man's presence, the minimum requirements met for the Folsom site still apply for future excavations. The primary requirement is a human skeleton, or an assemblage of artifacts that are clearly the work of man. Next, this evidence must lie in situ within undisturbed geological

deposits in order to clearly demonstrate the primary association of artifacts with stratigraphy. Lastly, the minimum age of the site must be demonstrable by primary association with fossils of known age or with material suitable for reliable isotopic age dating. (Haynes 1969: 714)

In fact, not until Tom Dillehay (1997) empirically proved in *Monte Verde: A Late Pleistocene Settlement in Chile 1: Paleoenvironment and Site Context* that people had been in the Americas before the Clovis era with the excavation and dating of the Monte Verde site in southern Chile did any of the other sites in the Americas, such as Meadowcroft, Cactus Hill, and Pedra Furada become seriously considered.

Evidence from Linguists and Molecular Anthropologists

Along with archaeologists, linguists also entered the debate concerning the peopling of the Americas, either siding with those who favored a Bering land bridge model or with those who came to be known as "early entry" modelers and who argued for a pre-Clovis peopling of the Americas. In contributing to the debate, linguists analyzed Native American languages and argued that Native Americans had been in the Americas for over 12,000 years based on interlanguage separation and diversity. Furthermore, based on the location and density of different languages found throughout the Americas, according to Lyle Campbell (1997) in *American Indian Languages*, linguists argued that the Americas were colonized in a west-to-east, north-to-south gradient. Molecular anthropologists also contributed to the debate by examining the frequencies of certain genetic traits in both Native Americans and indigenous Asian populations. The molecular anthropological data indicated that Native Americans came from several different geographic locations within Asia and that the various genetic lineages had different time depths. As a result of these multiple lines of evidence, the debate shifted from favoring a Bering land bridge model of colonization to one that took into account multiple pathways and modes of transportation in the colonization process. One of the different modes and pathways considered was the Northwest Coast route via watercraft and glacial refugias. The current empirical evidence conclusively proves that Native American peoples came to the Americas via several pathways, one along the Northwest Coast route through the use of watercraft, and another via the Bering land bridge and terrestrial means.

Early Debate Camps

Early debates concerning the peopling of the Americas, as a result, could be placed into two camps: those favoring the "early entry" model or those favoring the Bering land bridge model. The Bering land bridge was a landmass that resulted from sea levels lowering during the Pleistocene epoch, and evidence

indicates that this land bridge remained generally ice free even during continental glaciations of the Pleistocene. It stretched from the present-day Kolyma River in Siberia to the Mackenzie River in Canada, a distance of over 3,200 km in length. The Bering Strait was located at the center of the land bridge, which was bound on either side by the Chukchi and Bering seas. During glacial periods, the worldwide lowering of the sea level caused the continental shelf beneath the Bering Strait to be exposed as dry land, resulting in an intercontinental connection that stretched 1,600 km north to south.

Paleontologists, according to Richard Morlan (2003) in "Current Perspectives on the Pleistocene Archaeology of Eastern Beringia," believe that the Bering land bridge was a major pathway for faunal exchanges between the hemispheres, as well as an important evolutionary center on its own, resulting in the area's distinctive regional fauna. Because the Bering land bridge was so extensive, it was originally argued that early Native Americans must have followed caribou and other fauna across this land bridge. However, because the Laurentide and Cordilleran ice sheets covered almost all of interior Canada during the Pleistocene, it was also hypothesized that early Native Americans were trapped in interior Alaska until the ice sheets melted at the end of the Pleistocene. This picture, however, is no longer tenable based on recent empirical data. Rather, the evidence now indicates that early Native Americans also used watercraft to hop from glacial refugia, nunataks, and coastal islands that remained ice free during the Pleistocene.

Supporting Evidence for Boat Arrivals

This new understanding is supported by several independent lines of evidence and conclusively resolves the debate. Evidence indicating that early Native Americans traveled along the Northwest Coast involves several different empirical lines of evidence. For example, there is now substantial evidence indicating that a pervasive ice sheet did not cover the Northwest Coastal ranges, but rather during the Pleistocene, there were a series of valley and alpine glaciers, punctuated by ice-free refugias and nunataks. Similarly, evidence, as discussed by Julie Brigham-Grete and colleagues (2003) in "Chlorine-36 and 14C Chronology Support a Limited Last Glacial Maximum across Central Chukotka, Northeastern Siberia, and No Beringian Ice Sheet," indicates that Pleistocene glacial ice extent across Chukotka and northwest Alaska was noncontinuous, and most of central and western Beringia was arid. However, when global sea levels rose and inundated the Bering land bridge, much of the exposed Northwest Coast was also inundated, covering most of the evidence of these early migrants. N. Toth (1991), in "The Material Record," has suggested that if we assume a model for the colonization of the Americas that uses an ever-increasing population over time, the odds of documenting the very earliest evidence of human occupation are very slim.

Monte Verde

A site in modern-day Chile about 40 miles from the Pacific, Monte Verde was excavated in 1977 by American Tom Dillehay and Chilean Mario Pino. The site was home to about two dozen people, who lived in a tent-like structure divided into individual rooms by walls of animal hide. Each room was equipped with a clay brazier, and two communal hearths for tool making sat outside the tent. The many plant remains from over 100 miles away indicate that the Monte Verdians either traveled great distances to collect food and other supplies or had a trade relationship with some other group.

Further, there is no need to think of human migration as a specific event. According to D. J. Meltzer (1989), in "Why Don't We Know When the First People Came to North America?," humans may have populated the Americas in small numbers, or migratory dribbles, over long periods of time, an assumption that accords with the Northwest Coast route model. Complicating this picture is the fact that some of these small groups of early migrants could have been genetically swamped by later groups, exterminated by warfare, or eliminated by the introduction of disease. Groups may have been too small to be viable or may have died out because they were unable to adapt to new environments. If the earliest migrants were few in number and used artifacts primarily manufactured from perishable organic material that survived for only a short time, the evidence of their presence would be extremely difficult to detect in the archaeological record. According to James Dixon (1999) in *Bones, Boats, and Bison*, this would be even more difficult if these early people lacked what archaeologists consider to be diagnostic artifact types, such as fluted stone projectile points.

Despite these initial limitations, several scholars have argued in favor of this model over time. The most influential proponent of the Northwest Coast route has been Knut Fladmark (1979). He argued in "Routes" that with the use of watercraft, people gradually colonized unglaciated refugia and areas along the continental shelf exposed by the lower sea level. Fladmark's hypothesis received additional support from Ruth Gruhn (1988), in "Linguistic Evidence in Support of the Coastal Route of Earliest Entry into the New World," and R. A. Rogers (1985), in "Wisconsinan Glaciation and the Dispersal of Native Ethnic Groups in North America," who pointed out that the greatest diversity of Native American languages occurs along the West Coast of the Americas, suggesting that this region has been settled the longest.

Biological, Archaeological, and Linguistic Evidence

Currently there are three independent lines of evidence—biological, archaeological, and linguistic—that conclusively support the hypothesis that Native

Americans came to North and South America by watercraft as well as by land bridge.

The biological evidence consists primarily of dental and skeletal remains, as well as mitochondrial deoxyribonucleic acid (mtDNA) and Y chromosome genetic data. Dental evidence, according to Christy G. Turner II (1994) in "Relating Eurasian and Native American Populations through Dental Morphology," indicates that early Native Americans migrated into the Americas from northern Asia sometime around 20,000 years ago. Based on about 20 dental traits, such as the shape of the tooth crowns and the number of tooth roots, Turner defined an overall dental pattern called sinodonty. This distinctive dental pattern is shared among Native Americans and people from northeast Asia. Another, less complex dental pattern called sundadonty, is shared among people of Southeast Asia and occurs in some early Native American skeletons. Turner concludes that because there has been less dental evolution in North and South America, the Americas have been occupied for less time than Asia, and that widespread sinodonty demonstrates a northeast Asian origin for Native Americans. This evidence also indicates that there were several migratory waves of people from north Asia into the Americas, one along the Northwest Coast of North America via watercraft and then another one later that moved through the Bering land bridge and into the plains of America. Early skeletons found along the coast of North America and its tributaries exhibit sinodontal traits, such as the Kennewick skeleton found on the Columbia River and dating to 9,000 years ago, while those found more inland exhibit sundadontal traits.

Craniometric data, while not providing any chronological dates to tie down these migrations, also indicate that Native Americans arrived in the Americas at the end of the Late Pleistocene. The craniometric data provide further evidence that the Bering land bridge was not the only means by which Native Americans migrated to the Americas. For example, findings from Santana do Riacho in Brazil exhibit strong morphological affinities with present-day Australians and Africans, yet show little resemblance to contemporary northern Asians and Native Americans. Walter Neves and colleagues (2003), in "Early Holocene Human Skeletal Remains from Santana do Riacho, Brazil," argue that this indicates that two very distinct populations entered the Americas by the end of the Pleistocene and that the transition between the cranial morphology of the first migrants and the morphology of later Native Americans, which occurred around 8,000 to 9,000 years ago, was abrupt. Within these populations, long, local mircoevolutionary processes also mediated by selection and drift took place, accounting for the discrepancy between contemporary Native Americans and the first colonizers of the Americas. In fact, like the dental evidence, the craniometric data argue for the conclusion that an early migratory group came to the Americas by means of watercraft along the Northwest Coast route, quickly moving down the coast into Central and South America. Later migratory groups apparently originated in

central and northeast Asia, using both the Bering land bridge and the Northwest Coast, spreading inland as the Pleistocene glaciers retreated.

The genetic evidence, based on haplogroup frequencies of genetic markers found on both the mtDNA and the Y chromosome, corroborates this conclusion. The first Native Americans reached the Americas between 20,000 and 35,000 years ago from northern Asia or Siberia. These early migrants are hypothesized to have followed the Northwest Coast route until they were south of the glacial ice sheets, where they expanded into all continental regions and brought with them mtDNA haplogroups A–D and Y chromosome haplogroup P-M45a and Q-242/Q-M3 haplotypes. The molecular genetic evidence further indicates that a later migration entered the Americas, bringing mtDNA haplogroup X and Y chromosome haplogroups P-M45b, C-M130, and R1a1-M17, possibly using an interior route.

Another line of evidence arguing for such a model is archaeological. With the exception of the Diring Yuriakh site, which is not universally accepted by all scholars, no credible archaeological evidence of human presence earlier than the late Pleistocene has been discovered in north Asia. The oldest accepted sites in north Asia are known from the middle and lower Aldan River of Sakha (Ust-Mil 2 and the Ikhine sites), which date to around 35,000 years ago. The largest group of sites in north Asia thought to be tied to the first Americans lies in the Lena river basin, which contains a massive river system that drains the Central Siberian Plateau and higher upland areas and mountains to the south and east. With the exception of Dyuktai Cave on the Aldan River, these sites all represent open-air occupations located on low or medium river terraces near stream confluences. Localities on the Aldan River include Ust-Mil 2, Ikhine 1-2, Verkhne-Troitskaya, Ust-Timpton 1, Ezhantsy, and Tumulur. Several important sites from this period are also located along the Olekma River, including Kurung 2 and Leten Novyy 1.

Evidence further indicates that these north Asian first Americans most likely moved into the Lena basin from two concentrated areas: the Middle Amur River and the southern Primorye. Sites located in the Middle Amur area are found in the Zeya-Bureya Lowland and include one possible early locality along the Zeya River (Filimoshki) and several sites along the Amur River (Kumara 1-3). Sites in the southern Primorye occur on both sides of the Sikhote-Alin Mountains, which extend along the coast of the Sea of Japan. They include Geographical Society Cave on the Suchan River and several open-air localities situated along small rivers, most notably Osinovka and Gorbatka 3 on the Illistaya River, and Ustinovka 1 and Suvorovo 3 on the Zerkalnaya River.

All of this evidence argues that the early Native Americans focused their subsistence economy around rivers and the ocean and were familiar with watercraft and water travel. As such, it is highly logical that these early Native

Americans would have kept their subsistence economy as stable as possible as they expanded into new lands. Further, in relatively linear environmental zones, such as along coastal margins, colonization might be expected to be rapid, possibly resulting in a rapid migration down the North American coast and into Central and South America in conjunction with the use of watercraft. With colonization occurring along major environmental zones, it is reasonable to assume that different environmental regions of the Americas were colonized at different times and possibly at different rates of speed. For example, coastal zones may have been inhabited long before the interior plains or deserts, a conclusion supported by the archaeological evidence of the earliest known sites. The Northwest Coast model suggests an economy based on marine mammal hunting, saltwater fishing, shellfish gathering, and the use of watercraft. Thus, the Northwest Coast route would have proven to have similar subsistence resources as those familiar to early Native American migrants originating in north Asia and proven more attractive than an interior route. Specific archaeological sites supporting this model include On Your Knees Cave in southeast Alaska, Namu and Werner Bay in British Columbia, several sites on the Channel Islands of California, and Quebrada Jaguay, Quebrada Tacahuay, Quebrada Las Conchas, and Monte Verde along the western coast of South America. For example, according to James Dixon and colleagues (1997), in "Late Quaternary Regional Geoarchaeology of Southeast Alaska Karst," the data from On Your Knees Cave indicate that early Native Americans were experienced coastal navigators and adapted to the maritime environments of the Northwest Coast.

Perhaps some of the strongest evidence demonstrating that early Native Americans used watercraft to reach the Americas comes from the Channel Islands. These sites were never connected to the mainland during the last Ice Age, and at the time of earliest human occupation would have been separated by a 10-km-wide channel. There is also evidence, as discussed by Jon Erlandson (1994) in *Early Hunter-Gatherers of the California Coast* and Erlandson and colleagues (2006) in "Beads, Bifaces, and Boats" that these early Native Americans followed a maritime subsistence practice that included hook-and-line technology and nets made out of grass cordage.

The linguistic evidence also argues that the first Native Americans came from north Asia. This is based on phylogenetic language comparisons that indicate that Native American languages are more closely linked to north Asian languages than any other known languages in the world. Linguistic analysis, however, is unable to give any finer resolution to the question of where in north Asia the first Native Americans originated. Furthermore, the linguistic evidence also supports the argument that multiple pathways of entry were used by early Native American migrants. For example, there is much greater linguistic diversity in territories that drained into the Pacific (37 language roots) and along the Pacific coast (22 language roots) than that on the Atlantic coast (only 7 language roots). This

evidence indicates that the western coast of the Americas were inhabited longer than the rest of the continent, further supporting the arguments presented here.

Transoceanic Voyage Capability Evidence

Transoceanic voyages across the Atlantic or the Pacific have been considered as a possible means by which humans may have first colonized the Americas. Recent archaeological discoveries in Australia and western Polynesia indicate that ocean-going watercraft have been in existence for the past 35,000 to 40,000 years, and possibly for more than 100,000 years. The current evidence argues that early Native Americans colonized the Americas via a coastal route, moving inland at a later time. It is easier for people to exploit their current environmental zone, using their existing subsistence knowledge, than to move to adjacent foreign environmental zones. Thus people were likely to settle the coasts before moving very far inland. Current evidence indicates that the Northwest Coast route provided the environmental avenue essential for the initial human entry into the Americas. The coast formed part of a continuous marine coastal-intertidal ecosystem extending between northeastern Asia and northwestern North America and farther south to the Southern Hemisphere. It would have facilitated coastal navigation via boats and provided similar subsistence resources in a continuous ecological zone linking Asia and North America. With the use of watercraft the human population moved rapidly southward along the coastal-intertidal Pacific ecological zone.

Conclusion

Several independent lines of evidence conclusively argue that early Native Americans migrated to the Americas by means of several pathways. Although the old theory of the Bering land bridge still carries some weight, the empirical evidence indicates that this pathway was used only after the Pleistocene glaciers melted. Before this, early Native Americans followed the Northwest Coast route of entry into the Americas, using watercraft as they moved south along glacial refugia and ice-free islands.

References and Further Reading

Antevs, Ernst. "Review: The Holarctic Region." *Geographical Review* 28, 3 (1938): 520–21.

Bonnichsen, Robson, and Karen Turnmire. "An Introduction to the Peopling of the Americas." In R. Bonnichsen and K. L. Turnmire, eds., *Ice Age People in North America: Environments, Origins, and Adaptations* (pp. 1–26). Corvallis: Oregon State University Press, 1999.

Brigham-Grette, Julie, et al. "Chlorine-36 and 14C Chronology Support a Limited Last Glacial Maximum across Central Chukotka, Northeastern Siberia, and No Beringian Ice Sheet." *Quaternary Research* 59, 3 (2003): 386–98.

Campbell, Lyle. *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press, 1997.

Dillehay, Tom D. *Monte Verde: A Late Pleistocene Settlement in Chile 1: Paleoenvironment and Site Context*. Washington DC, Smithsonian Institution Press, 1989.

Dillehay, Tom D. *Monte Verde: A Late Pleistocene Settlement in Chile 2: The Archaeological Context*. Washington DC, Smithsonian Institution Press, 1997.

Dixon, James. *Bones, Boats, and Bison: Archeology and the First Colonization of North America*. Albuquerque: University of New Mexico Press, 1999.

Dixon, James, et al. "Late Quaternary Regional Geoarchaeology of Southeast Alaska Karst: A Progress Report." *Geoarchaeology: An International Journal* 12 (1997): 689–712.

Erlandson, Jon. *Early Hunter-Gatherers of the California Coast*. New York: Plenum Press, 1994.

Erlandson, Jon, et al. "Beads, Bifaces, and Boats: An Early Maritime Adaptation on the South Coast of San Miguel Island, California." *American Anthropologist* 107, 4 (2006): 677–83.

Fladmark, Knut R. "Routes: Alternative Migration Corridors for Early Man in North America." *American Antiquity* 44 (1979): 55–69.

Gruhn, Ruth. "Linguistic Evidence in Support of the Coastal Route of Earliest Entry into the New World." *Man* 23 (1988): 77–100.

Haynes, C. V. "The Earliest Americans." *Science* 166 (1969): 709–15.

Hopkins, David M. "Aspects of the Paleogeography of Beringia during the Late Pleistocene." In D. M. Hopkins, J. V. Matthews, C.E. Schweger, and S. B. Young, eds., *Paleoecology of Beringia* (pp. 3–28). New York: Academic Press, 1982.

Hopkins, David M. *The Bering Land Bridge*. Stanford, CA: Stanford University Press, 1967.

Johnston, W. A. "Quaternary Geology of North America in Relation to the Migration of Man." In D. Jenness, ed., *The American Aborigines, Their Origins and Antiquity* (pp. 9–46). Toronto: University of Toronto Press, 1933.

Jones, Peter N. *American Indian mtDNA, Y Chromosome Genetic Data, and the Peopling of North America*. Boulder, CO: Bauu Institute Press, 2004.

Krieger, A. D. "Comments on 'The Paleo-Indian Tradition in Eastern North America' by R. J. Mason." *Current Anthropology* 3, 3 (1962): 256–59.

Meltzer, D. J. "Why Don't We Know When the First People Came to North America?" *American Antiquity* 54 (1989): 471–90.

Morlan, Richard. "Current Perspectives on the Pleistocene Archaeology of Eastern Beringia." *Quaternary Research* 60, 1 (2003): 123–32.

Neves, Walter A., et al. "Early Holocene Human Skeletal Remains from Santana do Riacho, Brazil: Implications for the Settlement of the New World." *Journal of Human Evolution* 45, 1 (2003): 19–42.

Robinson, David M., and Elizabeth Pierce Blegen. "Archaeological News and Discussions." *American Journal of Archaeology* 37, 1 (1933): 117–59.

Rogers, R. A. "Wisconsinan Glaciation and the Dispersal of Native Ethnic Groups in North America." In T. C. Blackburn, ed., *Woman, Poet, Scientist: Essays in New World Anthropology Honoring Dr. Emma Louis Davis* (pp. 105–13). Los Altos, CA: Ballena Press, 1985.

Toth, N. "The Material Record." In T. D. Dillehay and D. J. Meltzer, eds., *The First Americans: Search and Research* (pp. 53–76). Boca Raton, FL: CRC Press, 1991.

Turner II, Christy G. "Relating Eurasian and Native American Populations through Dental Morphology." In R. Bonnichsen and D. G. Steele, eds., *Method and Theory for Investigating the Peopling of the Americas* (pp. 131–40). Corvallis, OR: Center for the Study of the First Americans, 1994.

CON

The question over the Bering Strait land bridge has been controversial and heavily debated ever since it was put forth. Some see it as racist, because it claims that Native Americans were never indigenous to the Americas, but initially came from elsewhere. Others have issues with time and dates. For that group, there needs to be clear, concise dates that the bridge was open. Such evidence proves elusive, and it seems that too many people worry about that question as the only question even in the face of a great body of evidence. Still others cannot believe that people dispersed throughout the American continents as rapidly as they did. To have done so, according to that group, people would have had to have arrived much earlier than the firm date range that many consider as being the only time the land bridge was not submerged.

There are many issues with these questions and arguments. First, it seems clear that the Bering Strait land bridge was open at different times. It was submerged for a period and then reappeared before disappearing again. The time frame available is flexible and varied. Furthermore, many people imagine the land bridge as simply an oversized footbridge, when it was in fact a great landmass with variations in geography and even in climate. The evidence is clear

Meandering river flats near the coast of the Bering Sea in the Bering Land Bridge National Preserve, Alaska. The preserve is a remnant of the land bridge that connected Asia with North America more than 13,000 years ago. (Greta Burkhart/National Park Service)

that the Bering Strait land bridge, also known as Beringia, was open around 15,000 to 18,000 years ago. However, it also appears that it was open prior to that as well. (O'Neill 2004)

Archaeological evidence suggests that the most active period of migration through Beringia started around 18,000 years ago. However, there is a strong evidence of several other periods where the ocean levels dropped enough for the land bridge to be exposed: 50,000 to 60,000 years ago, 40,000 to 45,000 years ago, 28,000 to 33,000 years ago, and 13,000 to 23,000 years ago. During the latter period, ocean levels were 300 feet lower than at present, and a low-lying landmass was exposed. At that time the land bridge covered most of the Chukchi Sea and stretched as far south as the Alaskan Peninsula in North America to the Gulf of Anadyr in Siberia. (Madsden 2004)

Origins of the Theory and Early Debates

During the first half of the past century, academia studied Native Americans in much greater detail than they ever had before. The disciplines of anthropology and archaeology emerged and grew in popularity as well. For the most part, these early scholars took a conservative point of view regarding the origins of Native Americans and felt they were relatively recent arrivals in the Americas. The question that remained was where they came from in the first place.

During the early decades of the 20th century, archaeologists established that modern man and his predecessors lived and evolved in Africa, Europe, and Asia amid four major ice ages from approximately 1 million years ago to 10,000 years ago. Technological advances improved with each succeeding ice age. Using fire, improved shelters, better and more powerful weaponry, and clothing increasingly suited to the environment, these people were able to survive the harsh conditions presented by the ice ages. According to the theory put forth during this era by scholars, people long removed from the fully modern man reached the far eastern portion of Asia. Those people, commonly referred to at the time as Mongoloids, migrated through Siberia and crossed into the Americas via the Bering Strait between 10,000 and 15,000 years ago. At that time, the Earth experienced an ice age and ocean levels dropped to allow a land bridge to develop that Mongoloid hunters used to follow animals of prey.

Thomas Jefferson wrote *Notes on the State of Virginia* in 1781. In it, he postulated the Bering Strait land bridge theory. He discussed an Indian burial mound he had located and excavated, and his work led him to believe that the Indians of America came from Asia through the far north. After Jefferson, the noted naturalist and explorer Alexander von Humboldt gave further credibility to the theory when he described what he saw as a striking resemblance between Native Americans and the "Mongol race." Later anthropologists such as Harvard's Ernest Albert Hooton agreed. He discussed the mongoloid characteristics he felt all Native Americans had, including dark, straight hair, dark eyes, medium brown skin, relative lack of facial and body hair, and wide cheek bones. Eskimos, he said, were the last migrants to arrive and the most mongoloid of all, with their smooth foreheads, marked eyelid folds, "infantile" nose, and yellow skin. Needless to say, many of these arguments are incredibly problematic, especially because of their simplicity and overgeneralizations. Indeed, not all Native Americans were alike, and they had, as they continue to have today, a great deal of variation in both appearance and lifestyle.

Another group of people that must be mentioned are those who felt that even 10,000 years ago was too recent of an arrival time for Native Americans from Asia. The leader of this group, Ales Hrdlicka, who was with the Smithsonian from 1909–1941, said that Native Americans arrived in the Americas no more than 3,000 years ago. He said that there was absolutely no possibility of an ice age or glacial age community, because the massive sheets of ice that covered much of North America had to recede and melt before habitation could occur. He felt that people arrived by boat. The Bering Strait is 56 miles across at its narrowest point, and he said people at that time could have easily traveled that distance by boat.

The majority of scholars who read Hrdlicka's work believed him. They believed his theory so much that it took over 30 years for a serious debate and dissent to come forth against his ideas. There were clearly many people who believed in the Bering Strait land bridge and a number who did not feel it was

possible. On the far end of the latter spectrum were people like Hrdlicka. However, as the 20th century progressed and scientific research and dating techniques became more advanced, it became increasingly clear that people lived in the Americas long before the 3,000-year-ago point. Evidence pointed to habitation much, much earlier (Adovasio 2002).

Evidence: Folsom and Clovis Points

The first skeleton discovered in North America that dated from the last ice age, which occurred between 10,000 and 70,000 years ago, was found in Minnesota in 1931. It appeared to be a drowning victim who had many similarities to modern Native Americans. Opponents of the age of the skeleton stated that it was simply a recent burial of a member of the Sioux tribe.

Perhaps a cowboy in New Mexico found the most important piece of evidence during those early years of debate in favor of the Bering Strait land bridge. George McJunkin was riding near the Cimarron River in 1908 in northeast New Mexico, near the town of Folsom. While looking for lost cattle, something caught his eye. When he rode over to the glistening whiteness that he originally noticed, he found a bleached skeleton. He dismounted and began digging around the skeleton. While the bones of lost cattle were a common sight in that area, the skeleton in question was buried 20 feet below the surface of the ground. As he dug, he uncovered several flint projectile points that were quite different from the arrowheads commonly found in that region. In fact, almost every aspect of the projectile, from its shape, markings, and size, was different. Furthermore, the bones were different, because they were much larger than typical cattle bones. It turned out that McJunkin had located the skeleton of a bison, a species that had been extinct for around 10,000 years. His find was proof that areas of the North American continent were in fact inhabited during the last ice age. The area where he discovered the skeleton, now commonly known as the "Folsom Site," was excavated and explored by archaeologists in 1926.

A discovery a few years later gave further credibility to the Folsom find and also provided more evidence that illustrated the diversity of the first inhabitants of North America. Again, it was amateurs, rather than members of academia, who made the find. Near Clovis, New Mexico, a man found a projectile point that was different from the Folsom projectiles in 1929. In 1932, a formal excavation of the area commenced. The Clovis Site yielded the skeletons of extinct horses, camels, and at least four mammoths. As with other Folsom sites discovered after 1908, several projectiles were found imbedded in the skeletons themselves. (Dixon 1993) This proved that man existed alongside these animals. The Clovis projectiles were different from their Folsom cousins. They were longer and heavier, which corresponded to the larger prey the Clovis-wielding hunters pursued, namely the wholly mammoth. They hunted larger prey and

Molecular Anthropology

Molecular anthropology examines DNA sequences to determine relationships between population groups, which is helpful in tracing migratory patterns and the discovery of common ancestors. Molecular anthropologists have determined a close common ancestor for humans and chimpanzees, for instance, and by studying mitochondrial DNA, have hypothesized "mitochondrial Eve," a 200,000-year-old African woman from whom all humans must have descended. Some tangential support for theories about prehistoric matriarchies comes from the discovery that prehistoric women traveled more than men did, which might be an indication of their social importance. Genetic evidence has also been used to indicate the date that the non-African populations diverged from the African one—about 50,000 years ago. These dates are constantly adjusted as new finds are made, providing molecular anthropologists with more data.

therefore required larger, more powerful weapons. Archaeologists also found other artifacts at the Clovis site, including a knife, hide scraper, and other various projectile points.

Evidence: Monte Verde

Bands of hunter-gatherers reached modern-day South America at least 11,000 years ago. They chased after a diversity of game, including the guanaco and the ground sloth, which are now extinct. One of the better-known sites in South America is Monte Verde. It is located on the banks of a small river approximately 9 miles from the coast of the Pacific Ocean in modern Chile. Preservation at the location is remarkable. The river drains a moist bog in a humid sub-Antarctic forest that has been in the same place since the end of the last ice age. Not only do stones and bone fragments survive from the original settlers, but wooden objects do as well (Fagan 1987).

Excavations of Monte Verde revealed many interesting discoveries. The north side of the river has many areas where activity appeared to have centered. Twelve wooden dwellings were uncovered that were joined together to form two rows. Several of the dwellings were furnished with clay-lined braziers, and two large outdoor fireplaces were located. There was also a wishbone-shaped wooden structure whose purpose is not yet known. Support poles were found near that structure, leading some to believe that there could have been sidewalls around the wishbone made of branches.

There appears to be evidence that the people of Monte Verde at least exploited the mammoth, even if they did not hunt it. It seems that they were scavengers, because although mammoth bones were located, no butchery sites were found. Large quantities of edible plants and fruits were discovered, as well

as tools. There were modified mammoth bones and branches that were curved and possibly used as handles.

Evidence: Eastern United States

There have been a number of interesting discoveries that point to human habitation east of the Mississippi River around 12,000 years ago. Dozens of individual projectile points have been found. That points toward widespread, yet scattered, occupation from approximately 12,000 years ago. However, researchers have found more concentrated locations. Once such discovery is the Meadowcroft Rockshelter. This prehistoric site is located on Cross Creek, a small tributary of the Ohio River, approximately 30 miles from Pittsburgh, Pennsylvania. In the late 1970s a team of archaeologists, botanists, and soil experts excavated the area. It is a superb location for a shelter. It is of good size—914 square feet—and the prevailing winds carry both smoke and insects away from the shelter during the summer months. Excellent sources of fresh water, including natural springs, are located nearby, and the entire region was once plentiful with large game.

A number of specimens were retrieved during subsequent excavations. These included charcoal remnants, hand-made blades, a lance-shaped projectile point, and a great deal of debris generated from flaking larger stones. Furthermore, as the team dug through layers of strata, they found at least 45 species of mammals and 68 species of birds, in addition to a variety of plants. The region was a good location for habitation, and it appears that those among the earliest people to travel to North America found the location good as well.

Evidence: Southern Florida

Archaeologists working underwater off Florida have found a great deal of evidence that shows habitation dating back to between 11,000 and 12,000 years ago. A number of human remains, artifacts, and the bones of animals long extinct were located in natural sinkholes. The evidence located brought scholars to the realization that life in that part of the continent was quite different from what they previously thought, as well as being much different from what it is today. The evidence revealed that Florida was cooler and drier than it is today, and the use of the sinkhole areas by prehistoric peoples was ingenious.

At Warm Mineral Springs, located in modern Sarasota County, a human burial was located on a ledge more than 40 feet below the current water level. The body had been deposited into a grave that had subsequently been submerged under water. A shell hook was located with the skeleton that was dated to be approximately 10,300 years old. Again, it is clear that people were able to populate the Americas quite rapidly once they crossed Beringia.

Evidence: Sustainability of Life on Beringia

The climate of Beringia was incredibly harsh. It was one of the coldest places on Earth, and at the height of the Ice Age would have been even colder. It was also dry. On the surface it seems that Beringia was uninhabitable. However, many pieces of evidence were discovered that proved the opposite to be true.

Beringia was not necessarily a barren landmass. Both land-based studies and deep-sea cores show a varied environment that contained a number of rivers and an array of vegetation, although they had shallow roots systems, which prevented predictability, and periods of temperature changes. Nevertheless, Beringia was an arctic environment with no trees and in many ways was similar to the tundra seen in Alaska today. It should be noted that some scientists believe that the prevailing view of Beringia is a bit too pessimistic. In reality, they feel the land was much more diverse, more of a steppe than tundra, where a variety of animals lived and reproduced. Furthermore, in the center and south there were swamplands and shallow ponds where the environment became grass-dominated wetlands. Large rivers, including the Yukon, carried glacial melt water, which produced flood plains.

In the late 20th century a 400-square-mile area of Beringia, previously hidden under volcanic ash, was discovered. Found within it were samples of grasses, mosses, and other varieties of ground cover that appeared in a nearly continuous pattern. However, the root system was extremely shallow, which proved that the plants of Beringia had no long-term stability. There was also evidence of climatic extremes and strong winds. Beringia most likely had a low population density, and animals would have found the region to be most inhospitable as well.

One of the difficulties in studying the Bering Strait land bridge theory today is the fact that the land in question is under some of the coldest, roughest waters imaginable. Many important finds were unearthed on the shores of the Bering Strait. Furthermore, a number of specimens of large mammoth bones were recovered from the ocean floor. Luckily for modern researchers, the extreme cold of the region allows for ancient specimens to be found in near perfect condition. Occasionally evidence turns up in ancient animal burrows. One such find indicated the presence of both cottonwood trees and aspen trees on Beringia.

The study of Beringia vegetation was first made possible in 1916 by a Swedish botanist names Lennart von Post. He discovered that fossilized pollen grains that were found in Scandinavian bogs and marshes could be used to reconstruct patterns of vegetation during the past Ice Age over periods stretching thousands of years. Von Post discovered great vegetational variations in Scandinavia toward the end of the last Ice Age. His studies pioneered the development of a looking glass into Beringia climatic and vegetational history.

Dozens of core borings were taken from lakebeds, swamps, mud, and a variety of other deposits of Ice Age material. The findings were impressive, as

myriad plant and animal life was found to have possibly, and some say likely, existed in Beringia. The interpretations run the spectrum from those who feel that there is not enough clear, concrete, and unequivocal evidence to prove anything about the climate and vegetation of Beringia, to those who feel that Beringia was in fact a distinct ice age subcontinent.

Many scholars urge caution when too generous a picture of Beringia is presented. In the 1960s botanist Paul Colinvaux felt that there was a possibility that the southern region of Beringia had been warmed by the Japanese current and was consequently more hospitable. However, his findings proved that the southern region was just as harsh as the northern region, and he became a leader of the view that Beringia was a mostly unproductive place. The food resources there were limited and unpredictable at best. As opposed to those who feel Beringia had a variety of landscapes, tundra was scarce and ice prevailed. However, none of this refutes the evidence that humans could and were moving across from Asia (Goodman 1981).

Argument against Boat Crossing

Proponents of the theory that humans entered the Americas much earlier than previous thought possible, perhaps 35,000 to 40,000 years ago, run into a substantial hurdle. Migrants and hunters who reached the far eastern reaches of Asia would have most likely seen nothing but water. To cross the Bering Strait, they would have needed some sort of boat. The weather in that region has always been unpredictable, and the rough seas would have made the 50-some-mile crossing nearly impossible with the watercraft available at the time. Scholars simply do not know how sophisticated the boats were. Inhabitants of the world tens of thousands of years ago could likely have known that animal skins stretched around a frame could make a boat, but not one capable of crossing the Bering Strait. Modern inhabitants of the Arctic hunt sea mammals, but do so cautiously. They rarely venture far from shore, and it is likely their predecessors exercised the same types of caution.

Other issues persist as well. The majority of the scholars who claim people inhabited the Americas up to 25,000 years ago generally talk of incredibly simple, basic people who had basic hunting, fishing, and gathering lifestyles. They also claim they had little knowledge of tool use. However, the amount of evidence to the contrary is staggering. Most archaeologists feel confident that the first settlers of the Americas were fairly sophisticated and highly adaptable people.

Conclusion

Anthropologists posit the Bering Strait theory as the way ancestors of all American Indians immigrated into North America. There is archaeological, biological, and

geologic evidence to support this, although Native American creation stories submit other ideas, and neither the Bering Strait theory nor creation stories can be proven conclusively. The biological evidence that supports the Bering Strait theory includes numerous physical similarities between Native Americans and eastern Asian populations, including skeletal features, hair, teeth, coloration, and the lack of body hair. Linguistic evidence also exists.

The largest debate over Beringia today is over the question of timing. Did the first people arrive more than or less than 15,000 years ago? Evidence supports both schools, although the less than 15,000 years ago school has more concrete evidence. In the early 20th century bison bones were discovered near Folsom, New Mexico, which were around 10,000 years old. Spear points were also discovered there, which proved that humans had been there at the same time. Other evidence shows a human presence at least 15,000 years ago, while a recent find in Pennsylvania suggests the possibility of a much older first settlement date. It is conclusive that people migrated from Asia into North America during the last Ice Age. The question that remains is whether they were the first inhabitants or not. The evidence at present points conclusively to the fact that they indeed were.

References and Further Reading

Adovasio, J. M. *The First Americans: In Pursuit of Archaeology's Greatest Mystery*. New York: Random House, 2002.

Dixon, E. James. *Quest for the Origins of the First Americans*. Albuquerque: University of New Mexico Press, 1993.

Fagan, Brian M. *The Great Journey: The Peopling of Ancient America*. New York: Thames and Hudson, 1987.

Fiedel, Stuart J. *Prehistory of the Americas*. New York: Cambridge University Press, 1987.

Goodman, Jeffrey. *American Genesis: The American Indian and the Origins of Modern Man*. New York: Summit, 1981.

Madsen, D. B., editor. *Entering America: Northeast Asia and Beringia before the Last Glacial Maximum*. Salt Lake City: University of Utah Press, 2004.

O'Neill, Dan. *The Last Giant of Beringia: The Mystery of the Bering Land Bridge*. Boulder, CO: Westview, 2004.

10

The ancient Egyptians used volunteers, not slaves, to build the pyramids.

PRO   Harald Haarmann
CON   Talaat Shehata

PRO

The pyramids of Egypt are more than landmarks, more than manifestations of the architectural know-how of the ancients. These structures impress by their monumental proportions, and the biggest, the pyramids of Giza, range among the Seven Wonders of the ancient world. The pyramids have spurred the imagination of the human mind for many generations, and the most fantastic speculations have been popularized about their origin, their builders, and their purpose.

For some, the pyramids of Egypt are a symbol of grandeur, of a great civilization that once existed and whose impulses are still felt in our modern world. For others, those constructions are a symbol of vanity, of the megalomania of omnipotent rulers who had monuments erected for celebrating their immortal memory. Those who live by the standards of our modern materialistic world have the least understanding for the cultural embedding of those monuments and the spirit of the civilization that created them.

In order to erect the pyramids, tens of thousands of workers were needed. In the 19th century, when archaeology was in its infancy, students of Egyptian history jumped to conclusions about the recruitment of the pharaoh's workforce. The easiest and seemingly conclusive way to explain the assemblage of such a huge workforce was that it must have been recruited from slaves. In the works of the theoretician of socialism, Karl Marx (1818–1883), the degradation of human beings in a state of slavery is decried and Egypt was scourged as one of the slaveholder societies of antiquity. There was no doubt in Marx's mind that the pyramids were the product of forced labor and, for him, these monuments were a symbol of inhumane conditions and cruelty.

In a way, it was common practice to associate the monumental architecture with slaves as laborers. That was the prevailing view of the time, the zeitgeist of the 19th century. With the rise of socialist ideology and its implementation in the socialist states of the 20th century, the popular image of the whip-driven slaves who built the pyramids persisted in the ideological perspective of history.

227

Egyptology and World War II

One of the leading Egyptologists was Iorwerth E. S. Edwards (1909–1996), whose book, *The Pyramids of Egypt* (1947), was one of the best-selling works on the pyramids.

Born in London, the son of a Persian scholar who worked at the British Museum in the department of oriental printed books and manuscripts, he went to Merchant Taylors' School and the University of Cambridge. He then studied Arabic before starting work at the British Museum where he worked for this next 40 years. During World War II he was posted to Egypt to help translate confidential documents for the British Foreign Office.

Because of the concerns about bombings during the war, most of the artifacts were hidden in the west part of the country. After the war, they were brought back to London, and Edwards had the task of sorting out the collection to prepare it again for the public. It was during this time that he wrote his book on the pyramids.

Also outside socialist ideology, this stereotype of the pyramid builders continued from generation to generation, and it inspired several Hollywood movies.

There are some famous monuments that were indeed erected by forced laborers, and these may have served as an analogy for explaining how the pyramids came into being. As we learn from historical sources, hundreds of thousands of forced laborers were engaged in the construction of the Great Wall, built between the 2nd century BCE and the 3rd century CE. This defensive structure runs for 5,000 km from east to west across northern China. During the period of construction, more than 100,000 workers died of exhaustion, and many of them were buried in the foundations.

In many cases, the workforce of monumental buildings remains anonymous. In one case at least, the forcefully recruited workforce of an amazing construction can be identified. One may ask people about the greatest monument of Jewish history, and most would look for it in the Near East, suggesting perhaps the lost temple of Solomon in Jerusalem. In fact, the greatest Jewish workforce that was ever assembled comprised the captives of war whom the Roman Emperor Vespasian (ruling from 69 to 79 CE) brought to Rome after the victory over Israel and the conquest of Jerusalem in 70 CE. The Jewish captives were forced into labor to erect the Coliseum in Rome, the biggest amphitheater of the Roman world, in the 70s CE.

A factor that is common to the Great Wall, the Coliseum, and the pyramids is the authority of a centralized state, which made the large-scale organization of a numerous workforce possible for such monumental enterprises. Since ancient China, the Roman Empire and the Egyptian kingdom waged wars with their neighbors, obtained slaves, and used them for different purposes and services, it seems conclusive, at first sight, to identify the workforce in all these cases as forced laborers.

It is noteworthy that the input of state organization is no prerequisite for the motivation to erect monumental architecture. The earliest monumental structures were erected by hunter-gatherers at Göbekli Tepe in eastern Turkey in the ninth millennium BCE, thousands of years earlier than the pyramids. Those bands of workers were not recruited by any political authority because elite power or dominance did not yet exist. Those who built the temple at Göbekli Tepe were volunteers whose work was inspired by their beliefs in forces of the supernatural, which were revered and celebrated.

Other great monuments of the Old World were constructed by people that cultivated plants, but these communities of early agriculturalists did not yet know the institution of a state. Those who erected the megalithic temples of Malta in the fourth and third millennia BCE and the stone circles of Stonehenge in Salisbury Plain (UK) in the third millennium BCE were sedentary people with agrarian lifeways, not slaves, who organized themselves as members of a voluntary workforce. The fruits of their labor have persisted throughout the ages up to the present.

Who Were the Pharaohs?

It is technically an anachronism to speak about the pyramids and the pharaohs' workforce. The title of pharaoh (based on Egyptian *per-aa* "the great house") was associated with the rulers of Egypt as late as the era of the New Kingdom, that is, from the 15th century BCE onward. The pyramids were built 1,000 years earlier. At that time the rulers of Egypt were addressed as *hemef* ("his majesty") or *nesw* ("king"). Therefore, the titulary "pharaoh" is not used in scientific literature referring to the Old and Middle Kingdoms in ancient Egypt.

Among the attributes that were in use for those kings who had the pyramids erected was *nebty* ("bull of the Two Ladies"). This was a metaphorical allusion to the two parts of the Nile Valley that started out as independent states and were united to become dynastic Egypt around 3050 BCE. These regions were personified as female figures: Nekhbet, symbolizing Upper Egypt (wearing a white crown in imagery), and Uadjet, the Lady of Lower Egypt (wearing a red crown). The pharaohs wore a double crown as a symbol of the unified state. In addition, the allusion to the "Two Ladies" also makes reference to the dimensions the king united in his persona—the cosmos and the land—both of which were personified as female figures in Egyptian mythology.

Society in ancient Egypt was strongly hierarchical, with distinct social and professional groups being assigned their particular roles in the conventionalized web of social conduct. The state of Egypt was theocratic because the ruler (i.e., the pharaoh) was perceived as of divine descendancy. Being the incarnation of Horus, a former sky god and offspring of Osiris and Isis, the king "was the absolute authority, the ruler for life, and the intermediary between the gods and mankind" (Brewer and Teeter 1999: 411). The goddess Isis is often sculptured

in her role as mother holding the Horus child. In the later tradition, the pharaoh was identified as the son of the sun god Ra. Under the auspices of the Egyptian system of sacral kingship, whatever decision the pharaoh made, its contents became law and was absolutely binding.

The absolutist state of the divine ruler did not only bring an accumulation of privileges to the royal family, it also stipulated ample responsibility of the central figure of Egyptian society for the well-being of his subjects. The pharaoh took the divine assignment as the guarantor of the continuity of life in the order prescribed by the gods. The images of royal figures (i.e., of the pharaoh, his wives, and children) are associated with one of the foundational symbols of Egyptian worldview, the ankh, which, in its shape, resembles the motif of the cross. For this reason, the ankh symbol is sometimes referred to as the "Egyptian cross."

According to Egyptian beliefs, the ankh symbolizes the principle of life. The rhythm of existence in this world (i.e., birth, growth, aging, death) and in afterlife depends on the proper operating of the powers of the ankh. A network of relationships based on trust arises. The pharaoh is entrusted, by his divine father, with the responsibility of surveillance of the right order of life. The subjects in the pharaoh's realm put their trust in the ruler that he will fulfill the divine task to keep up the right order in the world.

The ancient Egyptians imagined that the right order would be transmitted— in addition to proper education in Egyptian families—through a natural medium, the air. The patron of the right order (Egyptian *mayet*) was the goddess Ma'at who is sometimes depicted as winged. The right order the pharaoh represented was literally "breathed in" by his subjects.

Which Pyramids?

At this point of the inspection of the scenario of pyramid building and the organization of the workforce, an identification of the architectural complexes in question is needed. Pyramids in Egypt were built at different times, and the knowledge of the chronology of building periods facilitates assessments about the availability of slaves in greater numbers.

A pyramid is not just one clearly defined form of architecture since, in the history of ancient Egypt, different types of pyramids were erected. In the various structures that are subsumed under the category of pyramid, several phases of development can be discerned. The earliest pyramids were step pyramids, of the type called *mastaba*. The first *mastaba*, in six steps, was erected by king Zoser, second ruler of the 3rd dynasty (ca. 2686–2613 BCE). The first monument that can be defined as a "true" pyramid was built by Seneferu (Snefru), first ruler of the 4th dynasty (ca. 2613–2494 BCE), who founded a necropolis (precinct for royal tombs) at Dahshur. There, a pyramid with a slope of 60 degrees was constructed, which is the typical measure of a true pyramid.
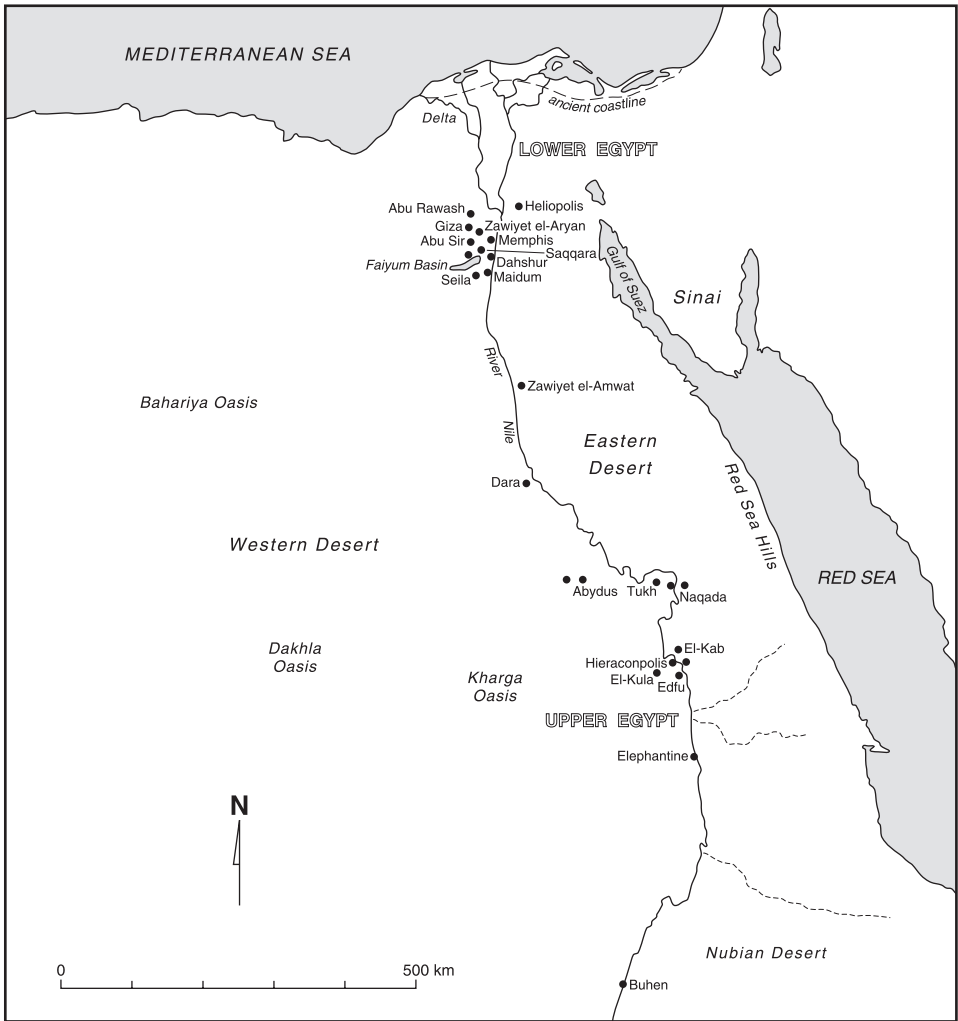
Egyptian archaeological sites.

Pyramidal structures were built in Egypt over a longer span of time (i.e., some 500 years), throughout the Old Kingdom, extending from the 3rd dynasty to the 6th dynasty and ending in 2181 BCE. Altogether 23 pyramids were built during that time, most of them in a relatively small area west of Memphis, at Giza, Zawijet el-Aryan, Abusir, Saqqara, and Dahshur. Two pyramids are outliers, located at some distance from the others at Abu Roash and Meydum (Maidum).

When we think of the famous Egyptian pyramids, we mean "the gigantic stone pyramids, the classic pyramids of popular imagination" (Lehner 1997: 640) at Giza, west of Memphis (south of Cairo). These pyramids that attract the most attention among visitors were erected during only three generations—roughly between 2590 and 2510 BCE, by Kh'ufu (Cheops), Khafre (Chephren), and Menkaure (Mycerinus). Kh'ufu's pyramid, the Great Pyramid, is the biggest

ever built, rising more than 146 meters from its base. It has been estimated that, for constructing Kh'ufu's pyramid, about 2.3 million stone blocks were used. The best-known view of the pyramids is perhaps that of Khafre's pyramid with the famous Sphinx in the foreground. The Sphinx is a hybrid figure, with a lion's body and a human head, featuring the image of King Khafre.

One can perceive the magnitude of these gigantic classic monuments when comparing them to other pyramids. The three most famous pyramids at Giza contain almost double the mass of stone material than all other pyramids of Egypt combined.

Pyramids were also built during the Middle Kingdom (ca. 2055–1795 BCE) and the New Kingdom (ca. 1550–1069 BCE). These structures, however, differed in size and function from the classic pyramids of the Old Kingdom. The later pyramids were much smaller and more like decorative upper parts of the main cult chamber of tombs, termed "pyramid tombs."

Were Slaves Available in Greater Numbers during the Era of the Old Kingdom?

The two predynastic state organizations in the Nile Valley that had emerged in the late fourth millennium BCE, Lower Egypt and Upper Egypt, were united around 3050 BCE to form pharaonic Egypt. The early dynastic history of the Old Kingdom, the period of the 1st and 2nd dynasties (ca. 3050–2700 BCE), was an era of consolidation of Egyptian traditions as we know them from later periods. During this time, the system of hieroglyphic signs, which had originated in the predynastic period, was further elaborated and conventionalized. The habit of furnishing the tombs with funerary statues was also perpetuated from predynastic times.

Settlements grew in size, and building activity increased. The typical genre of Egyptian architecture was still missing from the record of material culture in the early centuries of dynastic Egypt, and those were the monuments for which Egypt would become famous. The beginnings of monumental architecture lie with the 3rd dynasty in the 27th century BCE. State authority provided the organizational means to gather a workforce that had not yet been brought together earlier. There were ample human resources among the population of the Nile Valley to recruit the number of laborers that were needed for the big enterprise to build the pyramids. However, one social group in Egyptian hierarchical society was underpopulated at the time when the pyramids were built: the slaves.

Certainly, there were slaves in Egypt from the beginnings of pharaonic rule over the country. The slaves of the early days were assigned work in the households and in the royal precinct, but their numbers were far too small to provide the bulk of the workforce for the big public building projects. The reason for

this situation can be easily identified. The source for obtaining slaves in greater numbers was not yet activated in Egyptian politics of the 3rd dynasty.

The source that provided Egypt with greater numbers of slaves was war with its neighbors. Captives of war might have been brought to Egypt, as slaves, during the troubled years of the first intermediate period (22nd century BCE), but that was almost 500 years after the construction of the pyramids at Giza. It would take another 500 years before slaves in greater number would be put to work in the Nile Valley.

Slaves in greater numbers—either as captives of war or forcefully recruited workers in Egyptian-occupied foreign territory—were obtained by the Egyptians from two directions: the south (i.e., Nubia) and from the north (i.e., the Near Eastern region as far north as Syria).

Nubia and Its Human Resource for the Pharaoh's Workforce (Nubian Slaves)

Nubia in the south was the first foreign region that the rulers of Egypt tried to obtain. The first victory of the Egyptian army over Nubia is recorded for a ruler of the Middle Kingdom, Senusret I (Sesostris I), king of the 12th dynasty, in the 19th century BCE. The oldest pictures that can serve as evidence for the transfer of slaves from foreign territory into Egypt are reliefs on a stela of that period, showing rows of bound captives of war. That stela was erected at Buhen (a site at the Second Cataract) to commemorate the victory of the Egyptians over the Nubians.

At times, the relationship between Egyptians and Nubians was friendly; at other times they waged wars against each other. During periods of peace, Nubians were allowed to establish settlements in Egypt (e.g., north of Aswan). At times of war, captives from Nubia were brought to Egypt and forcefully recruited for the pharaoh's workforce. In the 15th century BCE, during the reign of three pharaohs of the 18th dynasty (i.e., Tuthmose I, II, and III), the Egyptian army launched decisive offensives against Nubia, which ultimately resulted in the conquest of the entire Nubian territory. The occupation of Nubia brought

> a steady supply of human beings. The role of slavery in ancient Egypt is not well understood but there are enough references in the texts to prisoners and depictions of them in tomb paintings and sculptured reliefs and on memorial buildings to make it clear that considerable numbers were taken to Egypt. . . . It seems likely that much of the work of restoring forts and erection of new buildings as well as the uncongenial work of quarrying for stone was done by these prisoners. (Shinnie 1996: 82*f*.)

Such recruitments, well known from the Egypt of the New Kingdom, added to the pharaoh's workforce 1,000 years after the pyramids of the Old Kingdom had been built. In no way could Nubian forced laborers have participated in their construction.

The Near East and Its Human Resource in Service to the Pharaoh (Asiatic and Aegean Slaves)

During the era of the New Kingdom, the Egyptians exerted political influence in Palestine and, to some extent, in southern Syria, although the grip of Egyptian supremacy was never stable in those regions that were never systematically colonized by the Egyptians, unlike Nubia in the south. From the mid-second millennium BCE onward, smaller groups of nomads from the Sinai desert infiltrated the region of the Nile Delta, eventually establishing settlements. The time for recruiting slaves in the Near East came with the military campaigns of the Egyptian army against the alliance of the so-called Sea Peoples (from different regions of the eastern Mediterranean), who raided the coastal areas of the Near East and Egypt in the 13th and 12th centuries BCE. One ethnic group among these peoples became well known in Egyptian sources and in the Bible, the Philistines, who were later allowed to settle in Palestine (which they named) and in the Delta area.

The captives from the wars with the Sea Peoples were mostly forced into the Egyptian army. They participated in the Egyptian military campaigns in their own units and, after serving successfully, the veterans received land grants and continued to live in Egypt. The same holds true for captives from the western desert, the Libyans, who had participated in the alliance of the Sea Peoples and were defeated by the Egyptian pharaoh Merenptah (19th dynasty), successor of Ramesses II. For obvious chronological reasons, neither Philistines nor Libyans could have been among the pyramid builders of the Old Kingdom.

Even at times when slaves were available and could have been recruited for large-scale construction projects, this was not necessarily done. Illustrative of a situation of voluntary workmen (nonslaves) being engaged in the building project of a pyramid—though of a smaller size than the classic big ones—is the discovery of a village of workmen (Kahun) who built the pyramid at Lahun for Senusret II (Sesostris II), ruler of the 12th dynasty, in the 19th century BCE. The seasonal inhabitants of the village were free workmen, and

> the domestic wares, the workmen's tools, the agricultural equipment, weaving equipment, children's toys, the make-up and jewellery of the women, and the articles associated with their daily religious observances have all been discovered. (David 1986: 4)

Spiritual Incentives to Do Voluntary Work for the Ruler: The Evidence from Ancient Egyptian Mythology

Against the background of the history of the import of slaves to be used as forced laborers in service to the pharaoh, it becomes evident that slaves were no factor in the composition of the workforce mobilized to build the pyramids of the Old Kingdom. The workmen who were recruited for that enterprise came

from among the Egyptian population of the Nile Valley. They were voluntary workers. The question arises: What motivated tens of thousands of workers to do hard labor for a public construction project?

The answer does not lie in the conditions of the economic system of the Old Kingdom. The point was not how many resources in natural goods or wealth state authorities could mobilize to start the building projects. The search for the motivation that mobilized huge masses of laborers takes us to the realm of Egyptian worldview and belief systems, because it is here that one finds the clues to an understanding of the spirit governing the great enterprise.

It was believed that death was not the end of the life cycle but rather an intermediate stage on its continuum. The worldview of the ancient Egyptians was permeated with thoughts about life after physical death. The sum of the individual's actions in his or her worldly existence, that is, in the span of time between birth and death, became a measure of a person's status and well-being in the unlimited period of afterlife. If ankh symbolizes the principle of life, then *ba* is the symbol for the physical existence of a human being, and *ka* his or her immortal soul, which copies the identity of the deceased and enters the realm of afterlife.

Entering the realm of the dead was no automatic passage, but every deceased individual had to pass a test of validity. In an imaginary tribunal, the *ka* (soul) of the deceased was placed on the scales and measured against *mayet* ("truth; righteousness; right order"). The god of knowledge and learning, Thoth, presided over the final judgment and recorded the results. The god of the underworld, the
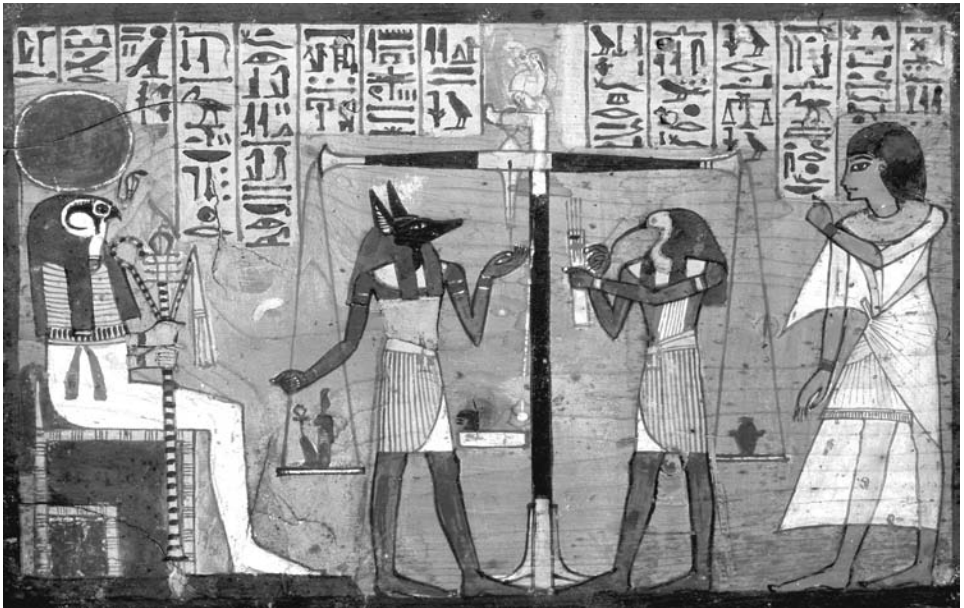
Anubis and Thoth weigh the heart of the deceased in the hall of judgment. Casket for Ushebtis, 21st dynasty (1085–950 BCE), Louvre, Paris. (Giraudon/Art Resource, NY)

shakal-headed Anubis, supervised the movements of the scales. In case the individual had accomplished good deeds during his or her lifetime, the *ka* would be heavy enough to tilt the scales in the individual's favor so he or she would be welcomed into the afterlife. The *ka* of those individuals who came to the tribunal with a bad record would be too light to reach a balance with *mayet*. Those individuals would be eternally damned and their souls would suffer in eternal unrest.

It is not surprising that the ancient Egyptians spent much of their time with preparations to ensure a decent afterlife for themselves. According to myth, the god Thoth gave human beings the gift of magic (*hekau* in Egyptian) to ward off evil spirits, which could put the good intentions of the king's subjects in jeopardy. Descriptions of social life and religious customs in ancient Egypt abound with accounts of preparations for afterlife.

Activities relating to preparations for the afterlife of a ruler were of the utmost priority and assumed a magnitude that surpassed any other enterprise in the Egyptian world. Erecting a pyramid comprised much more than merely constructing a burial site for a ruler.

> Whilst it is common to emphasize the mortuary character of pyramids and to see them primarily as tombs with temples ancillary to them, the way in which they were in fact organized and referred to suggests that the emphasis should be reversed, and they be regarded first and foremost as temples for the royal statues with a royal tomb attached to each, which, acting as a huge reliquary, gave enormous authority to what was, in essence, an ancestor cult and an important factor in the stability of government. (Kemp 1983: 85)

The pyramid was a symbol of sustainability of community life, both with respect to the institution of divine kingship (with the living king as guardian of the life principle) and to the memorial function in the ancestor cult (with the dead king as a revered divine ancestor). In addition, a pyramid was the materialized expression of people's devotion to the gods from whose ranks the pharaoh would come down to earth.

It is needless to emphasize that the king's request to construct a pyramid as a monumental manifestation of his status and role in society held the highest priority in a world where such a belief system functioned. Joining the pharaoh's workforce for such a majestic enterprise was not only a holy duty to serve Egyptian society but also a blessing that put the individual in the position to increase his or her personal merits for the passage to afterlife. Building a pyramid was the most sublime expression and manifestation of communal solidarity.

Organizing the Ruler's Voluntary Workforce

The construction of a pyramid requires diversified technical and mathematical know-how, ranging from the choice of the building site via the selection and

form of the stone material and the techniques to lift stone blocks to ever-higher levels. The command of separate technical skills does not suffice. For the grand scheme of a complete pyramid, there has to be a mastermind with the capacity to integrate all the partial aspects for the overarching view of the whole.

Some names of architects who built the pyramids are known. The architect of "Zoser's Step Pyramid, the first large-scale stone building in the world" (David 1986: 24), is the legendary Imhotep, the king's vizier. In the New Kingdom, Imhotep was identified as the son of the creator god Ptah. Later, he was revered as a god of healing and, in this role, compared with the Greek Asclepius. The architect of Kh'ufu's pyramid was Prince Hemiuen.

The construction of a pyramid requires many personnel who have specialized professional skills, among them masons, engineers, technicians, carpenters (who took care of the scaffolding), and others. In addition to those technical skills that were required for the construction of the monument itself, there were a host of other professional fields whose services were in demand. Workshops had to provide the many ropes that were used to move and lift the stones. For transportation, sleds were preferred, although the Egyptians knew wheeled vehicles. Specialized workshops produced the sleds according to the requirements of the technicians who were in charge of transportation of the stone blocks from the quarries to the building site. The sleds were drawn on specially prepared paths, paved with poles that were lubricated with animal fat to facilitate the movement of the vehicles. A fleet of barges had to be available for transporting stones over greater distances along the Nile.

Since the construction of the royal pyramid had the highest priority in building activities, the most skillful craftsmen from all over Egypt were summoned to accomplish this task. The specialized craftsmen, though, made up only a smaller portion of the entire workforce. The bulk were laborers, unskilled workers, who extracted stones in the quarries, engaged in the transport of the stone blocks to the building site and in their maneuvering into the right position, digging cavities for the foundations, amassing sand for the ramps to lift stones to a higher level, and so forth. For these tasks, many thousands of workers had to be mobilized and organized. In addition, overseers had to take care of vast supplies of foodstuff and its storage for the workers. Estimates of the number of workers range from 20,000 to some 100,000.

Among the social and professional groupings of ancient Egyptian society at the times of pyramid building there was only one group that was populous enough to allow the recruitment of masses of laborers from among them: the peasants who worked the fields in the Nile Valley. They made up more than 70 percent of the Egyptian population. The routine of horticultural and agricultural activities was seasonal, with intensive peaks during the periods of sowing and harvest.

The ecological conditions of the Nile Valley depended entirely on the fluctuations of the waters of the big river. The river would inundate at Aswan in late June and, during July, the fertile muddy slit would arrive and cover the

fields. Inundation would reach the area of Cairo in late September. Seeds were sown in fall. That was possible in October, once the waters from the river with its seasonal flooding had receded behind the banks. The river would reach its lowest level in the following year, in April.

Throughout the winter season, the peasants were available as a workforce that could focus on communal tasks without preoccupation for the crops. And this is the point of departure for the far-sighted management of human resources in service to the king. Perhaps Imhotep was not only the mastermind behind the architectural endeavor of the first large-scale stone building in the history of humankind for which there was no precedent, but he may have also been the first to make provisions for the gathering of the huge workforce that he would have to rely on. The mobilization of the peasants during the period when they were not needed in the fields was to become the greatest communal endeavor of Egypt's history.

Gathering a crowd of unskilled workers at the building site was only the first step. These workers had to be organized in smaller work gangs and assigned tasks according to the requirements of the work schedule. "Peasant farmers from the surrounding villages and provinces rotated in and out of a labor force organized into competing gangs with names such as 'friends of Kh'ufu' and 'drunkards of Menkaure'" (Hawass 1997).

The composition of work gangs was not stable but changed to optimize the efficiency of the team. What made a team efficient depended on the collective spirit of working together among individuals, on the one hand, and on the flexibility to cooperate with other teams in a wider range of coordinated building-activities, on the other. These gangs would be given assignments according to their skills in teamwork.

Although it had been assumed for some time that the pyramid builders of the Old Kingdom were peasants—as were those who built pyramids during the Middle Kingdom (see above for the workmen at Kahun)—it was as late as the early 1990s that settlements of those builders were discovered. As it sometimes occurs in the history of archaeology, the first settlement was discovered by accident. In April 1990, an American tourist who visited the pyramids at Giza was thrown off her horse that had stumbled on the remnant of a hitherto unknown mud brick wall. This event prompted excavations in the vicinity of the Great Pyramid, which lead to the unearthing of a vast area with the remainders of the houses of the seasonal pyramid builders.

From the distribution of graves in a nearby cemetery and the inscriptions found there, it became clear that those who commanded the workmen (called overseer of masonry, inspector of the craftsmen, or director of workers) were buried in one section and ordinary workers who died during the campaign in another. In the households, many items of daily life were found (e.g., pottery, tools, figurines from domestic altars). The settlement also harbored temples, and these sacred buildings

were mostly dedicated to Hathor, the goddess of motherhood and also of dance and music. Apparently the wives of the workers were engaged, other than in domestic affairs, in the service to the goddess, either as priestesses or as their aides.

The masses of laborers were trained to act in a cooperative spirit for their seasonal work at the site where the pyramid was erected. The same people were released to fulfill their duties as peasants and take care of work in the fields when it was the right time. Pyramids were built over many years, and the seasonal interruptions of labor became part of the regular schedule. After several seasons of building activity, the peasants had been conditioned to teamwork and, consequently, it would be unfair to continue to call them "unskilled," despite the lack of any formal professional training or specialization.

The labor for the divine ruler hardly brought any significant material gain for the peasants. And yet, one can imagine that those who participated in the construction project were apt for certain privileges, for instance, concerning the rationing of food supply for their families, the distribution of highly fertile patches of land, not to mention the special attention that these individuals must have received in their village communities. On top of the list of advantages of being a member in the king's workforce stood the reward of an accumulated merit that guaranteed access to a pleasant afterlife.

References and Further Reading

Aldred, Cyril. *Egypt to the End of the Old Kingdom*. London: Thames and Hudson, 1965.

Bard, Kathryn A., ed. *Encyclopedia of the Archaeology of Ancient Egypt*. London: Routledge, 1999.

Brewer, Douglas J. and Emily Teeter. *Egypt and the Egyptians*. New York: Cambridge University Press, 1999.

David, Rosalie. *The Pyramid Builders of Ancient Egypt. A Modern Investigation of Pharaoh's Workforce*. London: Routledge and Kegan Paul, 1986.

Hawass, Zahi. "The Discovery of the Tombs of the Pyramid Builders at Giza." 1997. Available at www.guardians.net/hawass/buildtomb.htm (accessed May 27, 2010).

Kemp, B. J. "Old Kingdom, Middle Kingdom, and Second Intermediate Period." In *Ancient Egypt: A Social History*, B. G. Trigger, B. J. Kemp, D. O'Connor and A. B. Lloyd. New York: Cambridge University Press, 1983.

Lehner, Mark. *The Complete Pyramids*. London: Thames and Hudson, 1997.

Pinch, Geraldine. *Egyptian Mythology: A Guide to the Gods, Goddesses, and Traditions*. Oxford: Oxford University Press, 2004.

Shinnie, P. L. *Ancient Nubia*. London: Kegan Paul, 1996.

Stierlin, Henri. *The Pharaohs Master-Builders*. Paris: Finest S.A./Éditions Pierre Terrail, 1995.

Trigger, B. G., B. J. Kemp, D. O'Connor, and A. B. Lloyd. *Ancient Egypt—A Social History*. Cambridge: Cambridge University Press, 1983.

Tyldesley, Joyce. "The Private Lives of the Pyramid-Builders." 2005. Available at http://www.bbc.co.uk/history/ancient/egyptians/pyramid_builders_01.shtml (accessed May 27, 2010).

CON

Most modern Egyptians and their foreign scholarly supporters hold strongly to a collective historical and cultural consciousness argument. The argument proceeds as follows: With a strong sense of local, national, and regional identity, most of ancient Egypt's populace volunteered when their free-time allowed (when each of their agricultural duties for their families and communities were completed on a seasonal cycle) in the most recent state and national projects being undertaken, in pyramid building. It was regarded by most Egyptians as a never-ending religious gesture of fidelity and goodwill to the god Amen-Ra and his deified representative on Earth, the pharaoh. By regularly contributing to the maintenance of stability and order to the afterlife and the material world which they inhabited on Earth, the ancient Egyptian's ultimate hope was that his or her individual, family, and communal life continue uninterrupted.

Such interruptions could occur, as Egyptians clearly recognized, in a heartbeat. Horrific change could unfold, through regular annual disasters, with the Nile River bursting its banks during its usual rise in the late summer months. During such events, the Nile River would overflow and destroy their harvest and, in most cases, the villages they inhabited. By volunteering, they were not merely conveying pride in their local and national identities, which they strongly possessed, but more significantly, they were looking out for their self-interest. By satisfying the god Amen-Ra and other lesser deities, they were contributing to the aversion of the abysmal state of famine, civil strife, and disunion that could ensue, as often related to them through their oral traditions by their parents and grandparents over the family hearth. The parents and grandparents had at one time or another in their lives witnessed the destructive consequences and imbalances. Their greatest fear was that such calamities would be revisited on their children, grandchildren, and great-grandchildren. The climate and natural environment clearly had a daily impact on Egyptians' lives. The local and national religious mission was centered on the primary belief that in any manner possible dictated by their immediate need to survive and continue to thrive, it was of the essence that they volunteer and contribute to their individual, communal, and national salvation.

To sum it up, ancient Egyptians' willingness to volunteer and contribute to the construction of their most current state and national project of pyramid building was a mission driven by pure love and equal self-interest, which was

tightly woven into their daily, local, communal, and national fabric. Egyptians believed the god Amen-Ra and the lesser deities would guarantee that, especially if that spirit of requisite devotion and acts performed in a ritual sense, by each Egyptian and his or her extended community, was undertaken in an altruistic and selfless manner. But were these volunteers the only workers in the construction of the pyramids?

By the beginning of the Pharaonic and Old Kingdom period (3100 BCE), boatbuilding had attained an admirable high standard. With the annual replenishing waters and fertile black silt earth that the ancient Egyptian farmers depended on to help settle Egypt's earliest agricultural communities, the harvest increased significantly in ways that their forefathers had never experienced. This provided more stability to the individual ancient Egyptian farmer's life and provided him with more free time that could be expended on national pyramid- and tomb-building projects throughout the nation. Having mastered the currents of the Nile, with the growing navigational skills exhibited by a few skilled Egyptian seafarers, the nation was prepared to challenge not only their surroundings, but time itself. Over the next 3,000 years, the task would prove Herculean, but in hindsight, over 5,000 years later, the ancient Egyptians appear to have succeeded in conquering time in ways that future civilizations and empires had fallen significantly short. The ancient Egyptians became the original gold standard bearers by which future civilizations and empires would measure themselves and others. That would not have surprised the ancient Egyptians, since they held to the very strong conviction that what they had accomplished was something that needed to be done. The awe by which Egypt's early civilization was held by the earliest Greeks and later Romans caused them to use it as a rich repository of ancient wisdom, as they evolved and progressed within their own civilizations. This practice and phenomena would continue in Western Europe, the United States,

---

Ancient Symbols: The Pyramids and Angkor Wat

The building of the pyramids and the great temples of Angkor, in modern-day Cambodia, built in medieval times, has caused controversy and debate. The Cambodian Royalist tradition celebrates King Suryavarman II, whose vision led to the building of Angkor Wat, the largest religious complex in the world. As a result, it appeared on the flag used by nationalist groups and by the country after independence in 1953.

In 1970 Jan Myrdal published a book on Angkor where he argued that the great temple and the other monuments were constructed by the peasants and slaves forced to toil, as had their compatriots in Egypt. This thinking was adopted by the Communist Khmer Rouge of Pol Pot when they came to power, and although they changed the flag, they left the silhouette of Angkor Wat in the middle of it. When they were overthrown by a rival Communist faction in 1979, the image of Angkor Wat remained on the flag.

and Asia, well into the current 21st century. This phenomenon would have deeply puzzled the ancient Egyptians. For, as earlier mentioned, they strongly felt that what they individually and collectively accomplished was something that needed to be done.

When it came to building the pyramids by 3100 BCE with the expertise and manpower available to them, the ancient Egyptians first had to settle on a site. The site would have to be in the direction in which the sun would set, for according to the ancient Egyptian's belief system, one would locate the underworld where the sun would set. The pyramids, therefore, would have had to have been built west of the Nile River. They would have built them on high ground, so that when the Nile's banks overflowed in the late summer months, the sites would remain untouched. With the heavy lifting and Herculean efforts that would be expended in moving the tons of limestone boulders from atop the ships moored at the river's banks, they would have located the site not too far from the banks. According to ancient Egyptian thinking, it would have needed to be completed in one simple motion—pull down the limestone boulders from the quarries, load it onto the boats, pull the boats at the river's edge, pull the boulders off the boats at the sites, and pull and place them on their permanent sites, after having chiseled the rocks into the measured diameters that they wished to complete. The coup de grace would be later transporting from the east bank of the Nile an abundant supply of fine-quality limestone from the Tura quarries, within a short distance of the capital, to encase their latest pyramid project with. It would be a genuine and magnificent offering to the god Amen-Ra.

The precise manner by which the huge finely chiseled and smoothed limestone boulders were lifted into place remains a mystery, but many theories abound. One method might have involved the use of workers within designated teams, which alternated regularly with other teams, in the heavy, backbreaking effort to raise the completed boulders into their permanent location on the site by the use of ramps. As each level was completed, the same effort would be exerted by the worker teams for the next level, after raising the ramp. Another method might have been using a circular path that was devised around each level of the pyramid, also by the use of ramps, as construction continued from the base of the pyramid to its peak. By the time the worker teams had reached the peak, they would remove the ramps, pulleys, and all manner of other contraptions used that allowed them to reach the pyramid's peak. Then they would apply the fine-quality limestone, from the top down, as they descended to the base of the pyramid. The final method suggested was by the early classical Greek historian Herodotus, and that was after his brief visit to Egypt. He mused, as he often did in many of his writings on Egypt, which often did not lend to accuracy or a neutral historical rendition of the facts, that Egyptian workers used levers to move their finished limestone blocks from one level to the next. But whatever method was actually used to place those blocks in their final resting place, it must have been an

awesome sight to behold by all those exhausted participants that ultimately uplifted their spirits. The new pyramid became, as many others before and numerous others after, a continuing genuine testament and reaffirmation to the omnipotence of their god Amen-Ra, their pharaoh, and their own nation's greatness.

The Ancient Egyptians Used Slaves to Build the Pyramids

From the initial survey, it seems as though some workers were volunteers, others may have been prisoners, and some may, indeed, have been slaves. But what was the definition of a slave in ancient Egypt? The initial idea of ancient Egyptians using slaves to build the pyramids was first referenced in the Old Testament, in the second book written by Moses, the book of Exodus. It states that "the people of Israel groaned under their bondage, and cried out for help, and their cry under bondage came up to God" (Ex. 2:23). But, also, earlier in Exodus, it made it clear, that "the Egyptians were in dread of the people of Israel" (Ex. 1:12) because their population continued to multiply within Egypt. The idea being that the new Pharaoh, or "new King over Egypt, who did not know Joseph" (Ex. 1:8), would attempt to use the Israelites as a useful distraction, from his attempt to consolidate his powers within Egypt and further to extort taxes from the entire Egyptian population, in order to build a stronger military force to fortify Egypt's borders and possibly extend its hegemony over its neighboring eastern, western, and southern neighbors. This would have continued a steady increase and influx of treasures with which he sought to glorify his reign and self. The Old Testament stipulated that Pharaoh therefore "set taskmasters over them [the Israelites] to afflict with heavy burdens; and they built for Pharaoh store cities, Pithom and Raamses" (Ex. 1:11). Also, Pharaoh and his taskmasters "made the people of Israel serve with rigor, and made their lives bitter with hard service, in mortar and brick, and in all kinds of work in the field; in all their work they made them serve with rigor" (Ex. 1:13–14).

If one were to look up the word "bondage," as described in Exodus, in the latest edition of *Webster's Dictionary*, it would read as an "involuntary personal servitude (as serfdom or slavery) [Middle English bonde 'peasant, serf,' from Old English bonda 'householder.'" Historians, cultural anthropologists, and other scholars have differed as to how future generations of believers and secularists should interpret and reinterpret the word "bondage." Was it the bondage of serfs, as the institution of serfdom existed during the European Middle Age? Or that of an indentured servant working the fields and households? Was it similar to pre-1860s slavery in the United States, where slaves did not have any legal rights within American society—for the simple fact, that American laws and the courts did not recognize the very existence of a slave to execute any matters in the courts or express their grievances against any injustices inflicted upon them by their owners or taskmasters? Was that similar to the set of circumstances with

the Israelites in Egypt, prior to the advent of Moses's pharaoh? The Old Testament states that the Israelites, from the time of Joseph to that of Moses, lived in Egypt for "four hundred and thirty years" (Ex. 12:40). That number of years is 41 years longer than the accepted estimate of when William Bradford and the *Mayflower* pilgrims first landed in what is presently Provincetown at Cape Cod, prior to their final landing in 1620 at Plymouth Rock, in what is currently the town of Plymouth in Massachusetts; also, 87 years less than when Christopher Columbus landed in the Bahamas, on October 12, 1492, in what was later renamed the town of San Salvador. The point being, that was a large undocumented swath of history where much transpired within the settled Jewish community Egypt between Joseph's time and that of Moses. Some historians, cultural anthropologists, and scholars question as to what kept the Jewish community at Goshen and other locales within Egypt during that period of time in Egypt, prior to the arrival of Moses's pharaoh on Egypt's political landscape. Did the story of the Exodus events color what had actually existed within the Jewish community and their relations with their Egyptian neighbors prior to the advent of the new pharaoh? After all, a few of those scholars ask, why would thousands of Egyptians willingly loan their silver and gold to their Jewish neighbors prior to their departure with Moses if there were not actually strongly held feelings of trust and amiability by the local Egyptian population toward their Jewish neighbors, let alone loaning the family heirlooms embodied in most of those treasures to a group of departing slaves? It doesn't seem to pass the simplest litmus test.

Pyramids served as royal tombs for Egyptian pharaohs beginning in the era of the Old Kingdom. The pyramids of Giza demonstrate a highly sophisticated level of engineering, and the method of their construction is still debated today. (Corel)

The conflict between Moses and Egypt's latest pharaoh was one within the royal court—a crisis within Egypt's leadership (Moses being perceived by Egypt's populace as a beloved Egyptian prince). Conflicts as such have always existed in Egypt's long history. So, according to the general Egyptian population, it was par for the course. It was of no concern to them. They suffered under the same oppressive policies and laws as their Jewish neighbors and other ethnic minorities. It all amounted to who held the scepter of Egypt's present and future. Some pharaohs, such as Joseph's pharaoh, ruled justly. Others, such as the adult Moses's pharaoh, ruled unjustly and tyrannically. It was a simple apple and oranges paradigm. The important matter was how Egypt's populace, and other ethnic minorities such as the Jewish population, accommodated themselves to the political, religious, economic, social, and cultural realities. What directly served their self-interest? This extensive 430 years, a defined time with its own rich and dense period of situational specificity regarding the Jewish presence in Egypt, has been a rarely researched period in Jewish, Egyptian, and world history. It should prove to be an extremely fertile and rewarding landscape and period for unbiased research by conscientious and dedicated historians and diverse scholars within the rich and vibrant traditions of Jewish, ancient Egyptian, and world religious, literary, social, and cultural histories.

Another instance of a reference to the use of slaves in building the pyramids was the classical Greek historian Herodotus. In his *The Histories*, he rails against the oppressive nature of pharaohs Kh'ufu (ca. 2589–2566 BCE) and Khafra (ca. 2558–2532 BCE), the pharaohs with the two largest of the three pyramids on the Giza plateau; the third pyramid belonging to Menkaure (ca. 2532–2503 BCE). He stated that it took 20 years for over 100,000 men slaves to build Kh'ufu's Great Pyramid. Herodotus's rendition was of the same nature, that it was not beneath Pharaoh Kh'ufu to use his daughter's sexual wiles in the local taverns to allow him to gather more money to finance the construction of the Great Pyramid. Much of his information appears to have been hearsay that was bantered around by high priests, who may have had their own individual scores to settle with the royal family. Modern classical historians have learned to take much of what Herodotus has written about his travels to Egypt and other foreign locales as private exaggerations, which often had no factual basis upon which to place his evidence. But no one can take it away from him that he did have a very fertile imagination. He talks about how he measured both pyramids by himself and reached the conclusion that Khafra's pyramid was 12 meters (40 feet) lower than that of his brother Kh'ufu. Anyone who has visited either of those pyramids and internalized the gigantic scale with which they were built would find it a stretch of one's imagination to see Herodotus scaling both pyramids in the sixth century BCE, pulling out his measuring tape, and measuring their heights in order to precisely determine that Kh'afra's pyramid was 12 meters shorter than Kh'ufu's. Also, another note he makes about Kh'ufu's son Menkaure, who Herodotus was

less critical of than he was with his predecessors, states that after Menkaure discovered from his physicians and oracle that he was dying and would most likely be dead within six years, he decided to party day and night. The idea, according to Herodotus, was that Menkaure, by doing so, would double the amount of time he would live on Earth (i.e., instead of 6 years, he'd be stretching it to 12 years, with no time factored in for any sleep, except for possibly very brief catnaps); he could go back to his physicians and oracle and accuse them of lying. Only Amen-Ra knows what would have become of them after that.

Another theory of slaves having built the pyramids alludes to competing prison gangs, who might have been promised by pharaoh's officials that those who proved they could pull their weight would either be pardoned or given an early release. Except this theory does not prove that slaves were actually used. These prison gangs could have been brought in to help the average Egyptian farmer who volunteered his or her services in building this latest national project. Possibly they could have been given specialized work by the taskmasters and kept separate from the rest of the average Egyptian nationals who were contributing their services.

But being a prisoner does not imply that you are a slave. As a prisoner, one's liberties are limited. In some cases, many of them would be regarded and possibly treated as indentured servants with a limited reward system. But, in no way does that imply that the person was a nonentity within the larger society. For all it was worth, the person still had some rights, or the officials would not be dangling the carrot of freedom (which a complete pardon or early release would have established) to the competing prison gangs. Also, slave labor, if it was used, could be used year-round. But, from archaeological evidence, it appears that this theory has gradually become discredited. The seasonal cycle period of work appears to have taken hold among most classical historians, archaeologists, and the scholarly community. As daily excavations in Egypt are regularly being unearthed, it appears slaves did not leave behind an immense amount of documents within different shaped houses or homes with interior decorations next to the pyramid sites at which they worked. This has included wills, veterinary, and genealogical papyri. The towns within which these workers lived while the pyramid-building projects were under way have revealed a multitextured and more profound insight into the daily lives of the Egyptian pyramid builder. In the end, the evidence is inconclusive but seems to support the idea that the pyramids were built by a mixed workforce of volunteers, prisoners, and, perhaps, slaves of one definition or another.

References and Further Reading

Bartlett, William Henry. *The Nile Boat: or, Glimpses of the Land of Egypt.* London: A. Hall, Virtue, 1850.

Brown, William. *The Carpenter's Assistant; Containing a Succinct Account of Egyptian, Grecian, and Roman Architecture*. New York: Livermore and Rudd, 1956.

Budge, Sir Ernest Alfred Thompson Wallis. *Cleopatra's Needles and other Egyptian Obelisks*. London: Religious Tract Society, 1926.

Budge, Sir Ernest Alfred Thompson Wallis. *From Fetish to God in Ancient Egypt*. London: Oxford University Press, 1934.

Clayton, Peter. *The Rediscovery of Egypt*. London: Thames and Hudson, 1982.

Clayton, Peter. *Chronicle of the Pharaohs*. London: Thames and Hudson, 1994.

Davis, Vivian and Renee Friedman. *Egypt Uncovered*. London: Stewart, Tabori, and Chang, 1998.

Feddon, Robin. *Egypt: Land of the Valley*. London: John Murray, 1977.

Geertz, Clifford. *The Interpretation of Cultures*. New York: Basic Books, 1973.

Grene, David, trans. *The History: Herodotus*. Chicago: University of Chicago Press, 1987.

Harris, J. R., ed. *The Legacy of Egypt*. Oxford: Oxford University Press, 1971.

Herodotus. *The Histories*, trans. George Rawlinson. London: Everyman, 1992.

Hogan, Patrick Colm. *The Mind and Its Stories: Narrative Universals and Human Emotions*. Cambridge: Cambridge University Press, 2003.

Kemp, Barry. *Ancient Egypt: Anatomy of a Civilization*. London: Routledge, 1989.

Knoop, Douglas, and Jones, G. P. *Freemasonry and the Growth of Natural Religion*. London: Butler and Tanner, 1942.

Lehner, Mark. *The Complete Pyramids*. London: Thames and Hudson, 1997.

Myrdal, Jan and Gun Kessle (photos). *Angkor: An Essay on Art and Imperialism*. New York: Pantheon Books, 1970.

Otten, Charlotte M. *Anthropology and Art: Readings in Cross-Cultural Aesthetics*. Garden City, NY: Natural History Press, 1971.

Tooby, John, and Leda Cosmides. "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments." *Ethology and Sociobiology* 2 (2003): 375–424.

11

Ancient Egyptian obelisks were raised by a hitherto undiscovered technology.

PRO   Talaat Shehata
CON   Patrick G. Zander

PRO

In order to understand the means by which ancient Egyptian obelisks were raised, it's important to gain insight into the purpose for their existence and benefits accrued by their presence. How were they constructed and then transported? This is a mystery that has interested and continues to interest archaeologists, Egyptologists, ancient historians, and the average layperson. Was there an undiscovered technology used by them, that 21st-century aficionados are not aware of? These are all important questions posed that need to be pursued to satisfying conclusions.

Besides the perfection and precision executed by the ancient Egyptians in most of their construction endeavors, other than the numerous pyramids and temple sites, as in the city of Karnak in southern Egypt, none were more imposing and elegant than their obelisks. Obelisks in the future would be imitated, stolen, and expropriated, such as in buildings in Washington, D.C., New York, London, Paris, Istanbul, and Italy. The rape of Egyptian antiquities and obelisks was so extensive that it's not fully known by most scholarly accounts how many priceless Egyptian pieces have left its antiquarian shores, as Rome's power increased throughout the Mediterranean. It is clear from early Roman accounts that the Emperor Hadrian fancied different-sized Egyptian sculptures. In the ruins of his villa in 1771, the head of an Egyptian female sphinx was found, among numerous other Egyptian statues, coin pieces, and amulets. The imposing granite found on ancient Egyptian obelisks, with their rich yet undecipherable hieroglyphic writings, peaked the curiosity and excitement of most Romans. Later, from 306 to 337, Constantine the Great, Rome's first Christian ruler, would plunder Egypt's obelisks in order to quench his insatiable taste for them. In time, an enormous obelisk that had been commissioned by Pharaoh Tuhtmosis III, around 1500 BCE, was torn from its base at Thebes and shipped to Constantinople, which is present-day Istanbul, to arrive a few years after Constantine's death. In 390, Emperor Theodosius I would command that the obelisk be permanently rested at the Hippodrome. With the advent of Islam, the Hagia Sophia Mosque would be built next to its imposing structure, where it continues to stand.

249

Egyptian obelisk at the Place de la Concorde in Paris, France. (iStockphoto)

With the arrival of European colonialism during the 18th and 19th centuries, only five obelisks would remain in Egypt, while at least 50 obelisks continue to sit in the capitals and main squares of these former imperial powers. The obelisks made their journeys from Egypt to Rome, where one stands at St. Peter's Square in front of the Vatican, where it was moved to from the great Circus Maximus in 1587. The obelisk had lain discarded and completely neglected for over a millennium. A papal order by Pope Sixtus V would have it transported to the main square in front of the Vatican, where it has remained as a constant and eternal fixture. This earliest Roman fascination with Egyptian antiquity can be best summarized in graffiti that was etched by an early Roman tourist onto an ancient Egyptian temple in Philae, in Alexandria, Egypt. The tourist alluded to how one will attain a long, prosperous, and happy life if he or she prayed in earnest to Egypt's goddess Isis. This also explains why, after Italy was eliminated by South Korea's soccer team in the quarterfinals of the World Cup in June 2002, Italian fans, shocked by the defeat, emotionally took to the streets and headed to the ancient Egyptian obelisk at the center of Rome's Piazza del Popolo to lament their individual and collective grief. Besides the heat of that summer day, which made the fountains in that square an alluring prospect in which to cool themselves, their individual and collective passion got caught up in a very revealing and subconscious manner, which many of the participants might have not been truly aware of at that historic moment. Feeling stunned, robbed, and devastated by the loss to South Korea, they brought their collective grief to the ancient Egyptian structure, which earlier Romans would have reflexively also done to alleviate the excruciating pain of their nation's loss. Like the tourist in an earlier age, they were hoping that God, or "the gods," would help them cope and shower them at that *shared* moment of national grief (one needs to *truly* understand the passion many Italian soccer fans exhibit for their local and national teams to appreciate this), through a long, deep, and *genuine* prayer, with a long, prosperous, and happy life—especially for Italy and its national soccer team. To gauge the intensity of emotion being exhibited at the time, a middle-aged woman

fan called out to a newly proclaimed Franciscan and Italian priest Padre Pio, "Oh Padre Pio, Padre Pio, pray for Italy!"

Other obelisks would find their way from Egypt to Florence, Paris, London, and New York. In Paris, one obelisk would be placed at La Place de La Concorde between 1831 and 1834; in London and New York, the two obelisks, better known as Cleopatra's needles, would occupy public areas provided for them in 1878, in the City of Westminster, England, and later in 1881, in New York City's Central Park. Some scholars have said that the phenomena of obelisk plundering and reerection in each of these world power's capitals was a strong indication, using current parlance, that all these European powers didn't possess the *innate* spiritual and psychic confidence essential to what constituted or established their credentials as imperial powers during their very brief ascendancy and ultimate collapse in world history. That is, these powers felt they needed (and continue to need, in order to establish their bygone *kudos* as individual global powers in world history) to expropriate ancient Egypt's *mojo* to bestow on themselves some sense of legitimacy. If that was actually and continues to be the case, then that would be a sad legacy, indeed, for modern and future history; and the rectification of such standing policies by these varying nations by the return of these obelisks, along with other plundered and pillaged Egyptian symbols and treasures to their rightful Egyptian owners, would be a first positive step. But, many Egyptologists, ancient historians, and scholars, possibly justifiably so and being cognizant of the cynical nature of nations' political and cultural leaders, do not seem to be holding their collective breaths. For many, most things ancient-Egyptian have become too tightly woven into the fabric of what constitutes and defines Western civilization's core *identity*.

The English term obelisk, known in French as *obelisque*, Latin as *obeliscus*, and Greek as *obeliskos*, derived from *obelos*, with "pointed pillar" as its meaning, was regarded as a four-sided straight pillar structure which would have a gilded minipyramid, better known as a *pyramidion*, at the apex of the structure. At its creation, an obelisk would have been a truly awe-inspiring sight. The gilded pyramidion, when struck by the sun's rays, would have projected a beautiful shining white light. As the ultimate representation of the ancient Egyptians' solar religious fixation, the obelisk would have taken on a complete life of its own, as the sunlight traveled down the four-sided smoothly sanded and polished structure and inundated the average Egyptian's sense of comfort and balance, *knowing* that the king of Egypt's god, Amun-Ra, was so concerned with their being and daily schedules, that through his sunrays, he was forever looking down on each person with light and warmth, from dawn to dusk. That was especially the case from 2686 to 2181 BCE, during the Old Kingdom.

It was during the Old Kingdom that pyramid and obelisk building reached their zenith. It was in old Heliopolis that the elderly priests of Egypt propagated the cult of Amun-Ra, which in time, after acquiring a very strong cult following,

would become the state's official religion. Pharaoh, in time, would be regarded as a divinity because of his recognition by most, if not all ancient Egyptians, as being the actual human embodiment of Amun-Ra, since he was his son. In time, the squat standing stone unearthed in old Heliopolis archaeological digs would be recognized by most Egyptologists and ancient historians as possibly the earliest prototype of what would evolve into a pyramid-building mania in Egyptian society. This squat stone was called a *benben* stone. The true meaning of that word in ancient Egyptian texts was in reference to a "rising," which would allow us to better understand why, when the obelisks were completed, the pyramidion at the apex of the structure was capped. Its meaning is taken from an ancient Egyptian pyramid text, which referred to a *benu*-bird, or in current slang a phoenix, who would become, with the arrival of the Book of the Dead, a gray heron bird with a straight and long beak and a crest that was double-feathered. According to the Book of the Dead, by being transformed into a *benu*-bird, one was creating a direct link to the god Amun-Ra and the sun. It would include the regular daily process of rebirth. Therefore, through this capping process, the ancient Egyptians were collectively giving the "green light" for Amun-Ra's sunrays to hit the pinnacle of the pyramidion at dawn and have his essence flood the internal and external essence of what was Egypt's daily society. That cap, at the pinnacle of the obelisk, would come to be called a *benbenet*, signifying the existence of a small pyramid structure at the very top of a long rising pyramid. It was to be fine, exquisite, and beautiful. It was built during the first Egyptian dynasty.

As if possibly being aware by instinct of the chaos existent geologically in Earth's primordial past, the *benben* stone mounds might have symbolized to the ancient Egyptians the *rise* of order from the goddess Nun's watery chaos. This order would be symbolized by the daily rise at dawn of the sun, which according to ancient Egyptian beliefs, occurred on the first day of creation. Ancient Egyptians viewed the *benben* stone mound as a representation of the god Atum's *petrified* semen. They regarded Atum as the god of Heliopolis. Not to be ever accused of lacking any deficiency in their imaginations, ancient Egyptians in their discovered earlier texts had attributed to this great god of Heliopolis a rather creative act of masturbation, which at its climactic conclusion, brought forth the representations of air, in the divinity of the lesser god Shu, and moisture, in the divinity of the lesser god Tefn't. Over time, the stone mounds were replicated in most Egyptian sun temples. These sun temples, by the fifth dynasty, would begin to play a significant role in royal funerary rituals and processions. These stone mounds, in time, would be directly identified with growing pyramid structures for the royal families' burial. They would also be directly related to the sun god Amun-Ra.

Obelisks were often built from red granite (with weathering, they took on a pinkish tone) carved out from one single piece of rock from the Aswan quarries in southern Egypt. An unfinished obelisk discovered in Aswan has provided information to archaeologists that other standing obelisks have never been quite

able to do. If that unfinished obelisk had been completed and then raised, most scientists agree that it would have stood as the tallest Egyptian obelisk in the world, at 41.75 meters, weighing an impressive 1,168 tons. The second tallest and heaviest would have been the Lateran obelisk, which was commissioned by Pharaoh Tuhtmosis III and expropriated by the Romans to stand near the ancient Roman palace in Laterano, which would later become a Papal residence. In the Middle Ages, the obelisk would occupy a prominent spot, which would be directly adjacent to the Basilica di San Giovanni—a building complex within the Apostolic Palace of the Lateran. Next to the unfinished obelisk in Aswan, Thutmosis III's obelisk in Laterno pales in comparison. The obelisk's height is 32.18 meters, and it weighs 455 tons. Pharaoh (Queen) Hatshepsut's obelisk in Karnak, at a distant third, measures 29.56 meters high and weighs 323 tons. The Vatican, New York, and London obelisks, all also commissioned by Thutmosis III, measure in height, 25.37 meters, 21.21 meters, and 20.88 meters, respectively; and 331 tons, 193 tons, and 193 tons in weight, respectively.

The unfinished Aswan obelisk provided insight into the stone-digging and quarry-lifting techniques that were adopted by the ancient Egyptians. It appears that dolerite hammers were used on all the sides of the agreed-upon single stone quarry piece that would be raised. After pounding and chipping away with the hammers, for long, hard, hot hours with temperatures regularly reaching 120 degrees Fahrenheit during the day for months on end, in a year's span, the most skilled among them would descend to the bottom of the carved out obelisk and continue to pound and chip away the granite, in unbelievable contortionist body positions. Ancient records have alluded to 5 to 10 percent of the quarry workers dying from the heat and exhaustion. So it would not be an intellectual reach to state that once the work was completed at the quarries, and later at the obelisk's final resting place, it would have been nothing short of a man-made miracle. The time and effort that would have been expended by these workers, on that earliest stage of the project alone, would have been excruciatingly backbreaking. That does not factor in that when that stage of the project was completed, the obelisk would then need to be lifted and transported to its permanent site. So, with the unfinished Aswan obelisk, all archaeological evidence indicates that it was abandoned by the project overseers after their discovery of a few potentially fatal cracks and fissures throughout the structure, especially in the middle of the emerging edifice. That must have been a very hard blow to most, if not all, the workers who had labored extensively for months on end.

At Hahtshepsut's obelisk at Karnak, the only remaining one of the four she had mounted at the temple, there's an inscription claiming that the obelisk was completed in seven months—that is, locating it, digging and chiseling it out, removing it, then transporting it by using sleds for the trip near the Nile banks, where upon arrival at its final destination, hieroglyphs were inscribed on all its four-sides, and then it was raised and finally stabilized on preconstructed turning

grooves permanently set at the base of its foundation. A magnificent feat, if that was based on actual facts. Except, as history has regularly taught us, some leaders, especially pharaohs such as Ramses II and Queen Hahtshepsut, among numerous others in Egypt and throughout the ancient world, felt the perverse need to embellish most, if not all, of the facts and tenaciously glorify multiple accomplishments, or often the lack thereof. The driving need to project preconstructed false images has lightheartedly proved to generations of scholars that the era they live in was similar to most eras that predated them, or evolved after them, in all civilizations that existed in world history. The facts often got overlooked by the original perpetrators, or lost, not so much in the magnificence of what was being accomplished, but rather, in *how* the story was being told. Even the ancient people had a ready small army of spinmeisters, at the behest of the "supreme" leader of the moment, to regularly construct and reconstruct the narrative at hand. That, since man's earliest beginnings, has emerged as an all too common human trait often witnessed in individual and collective fragile psyches and natures, especially, while straining to realize one's magnificent ambitions.

It is known that sleds that had been rubbed down heavily with animal lard and fat, as well as levers, ramps, a limitless supply of rope, and hundreds of workers, were used in the transport and raising of the obelisks. Multiple grooves on the sides of each of the obelisks were used by the workers to stabilize the multiton structure as it was being loaded onto the wooden barges that would transport them hundred of miles down the Nile River. The barges would be capable of withstanding the weight of each obelisk, since the outer skin would be so meticulously fitted, like a tight jigsaw puzzle, that when the thick airtight inner hull was bound by limitless amounts of rope, and the inner joints were secured, it would have been possible for the barge or vessel to withstand any overwhelming structural weight. That was accomplished without the incessant worry by the overseers and the workers as to whether the barge or large vessel in use would immediately capsize. The grooves on the side of the obelisks also helped support the structure as it was being unloaded onto the chosen site, and then later, being raised by four or five teams of men. Each team comprised 200 to 300 men. If the overseers felt that more men were required for the painstaking task of transporting or raising the obelisk, there was always an abundant supply of volunteer workers available, along with military conscripts, indentured servants, and imprisoned criminals who might have been promised amnesty or early release if pharaoh's overseers felt that they had put their individual and combined might into the project. Also, there were the ongoing contributions of prisoners of war, if they were available at the time. The participants would be spread out on all equidistant sides of the obelisk, pulling with their collective might on the large thick ropes tied around those precreated grooves. Strong cyprus wood beams were used to construct a large framework by which skilled workers first climbed all sides of each obelisks and sanded and polished them

down for weeks, who then turned them over to the skilled artisans who would meticulously etch in 1- to 1.5-inch deep images and hieroglyphs using smaller dolerite stones, then pieces of sandstone and quartzite. Finally, small hammer stones were used to chisel in the finest details. Once foremen and fellow workers completed their jobs, after a long wait, the skilled artisans handed the framework and individual obelisk back to the skilled workers. Their jobs would have required them to softly, once again, sand and polish down each of the obelisks, only this time, with meticulous care and skill. It would be polished to smooth perfection so that it could radiantly reflect the sun. Finally, when the job was completed on all sides of the obelisks, then the same large wood-beam framework that had supported the diverse skilled artisans and workers would be torn down, relocated, rebuilt, and if need be, bolstered anew. That would have allowed for it to endure the final stage of the project, which would have required for it to withstand the enormous pressure and weight of raising and stabilizing each of the obelisks. This tearing down and rebuilding of the wood-beam framework could have been an endless process of minor readjustments to the overall needs of the project.

Clearly, the project, from beginning to end, was not for the fainthearted. As for the contention that most, if not all of the obelisks, when completed were simply raised and hauled up preconstructed ramps and then slowly eased into prefilled sand bases, which would have constituted hundreds of tons of sand at the base that had hypothetically been gradually drained as the obelisk was being lifted, seems to be a creative idea and in some cases practical in scope, but, not proved beyond a shadow of a doubt. That this was actually the way the ancient Egyptians raised their obelisks and placed them permanently on a turning groove at the base of the foundations is unproven. The final trick was that each of these obelisks would support itself, mind you, into *eternity*, by standing on its own weight. Despite most Egyptologists' skepticism concerning the validity of the use of "sandpits" for the lowering of a 400-ton or more obelisk, they accept the fact that the use of sand was common at ancient temples and palatial construction sites throughout Egypt. Tomb shafts were known to have had sand poured down them, while large sarcophagi were eased down them. Ancient documents and hieroglyphic carvings have alluded to sand use in raising some smaller-sized obelisks. An ancient papyrus, over 300 years old, has one scribe comically goading a fellow Egyptian as to how sand needed to be removed from under one large monument to make room for another—that their pharaoh might be pleased. Whether that monument was a large statue, sculpture, or even an obelisk was not quite clear from the information on the papyrus. All the same, most Egyptologists, archaeologists, and other scholars strongly believe that sand was used for many construction projects, especially at ancient Egyptian temple sites. The primary puzzle for these scholars is how extensive its use was, especially when it came to raising into a vertical position a 400- or 500-ton or more single structure, especially a well-carved and proportioned obelisk. Raising a 30- or 40-ton obelisk

---

Experimental Archaeology

Experimental archaeology is the art and science of reconstructing ancient or prehistoric processes in order to inform a hypothesis. For instance, Thor Heyerdahl famously built his balsa raft, the *Kon-Tiki*, in order to sail from Peru to Polynesia—just as he theorized the first Polynesians had done. Proving that it can be done doesn't prove that it was done, but it at least addresses some of a theory's detractors. Attempts have also been made to replicate Damascus steel, Greek fire, and early medicines, the compositions of which are unknown because they were kept secret or because the ancient terms for their components are a mystery to us now.

The form of experimental archaeology best known to students is probably flint-knapping, which can be done by anyone with a few pieces of stone (and eye protection, however inauthentic it may be).

---

is quite a different undertaking than having to raise a 400- or 500-pound obelisk. What ancient technology, if any, was applied to it?

The use of large sandpits at the resting base of the obelisks might have been one of many approaches used by the Egyptians. But, with more diligent excavations and research, hopefully, in time, and much sooner rather than later, archaeologists, Egyptologists, and ancient historians will discover the final few pieces that will complete the puzzle and narrow the possibilities. That leaves us with possibly yet another creative means used besides the sandpits: the possibility of the use by the ancient Egyptians of an as of yet undiscovered ancient technology in their regular practice of raising their obelisks onto their permanent foundations. The idea of an undiscovered technology should be regarded just as valid as any other earlier possibilities considered. It's clear from the facts made available to Egyptologists and other scholars that two very important hitherto "undiscovered" technologies have been discovered in the process of obelisk building in ancient Egypt. First are the *turning grooves* discovered at the base of all the foundations of the stolen and fallen obelisks throughout Egypt. Each of these obelisks had to have turning grooves in place *prior* to being permanently placed on their foundations. Second, and possibly much more importantly, the ancient Egyptian worker had to bring with him an unquenchable passion and unconquerable spirit of self-sacrifice. Having mastered the arts and architecture, they appeared to have developed very keen, curious, and searching eyes. They always attempted to operate fluidly within the natural world that surrounded them, imposing their collective supreme wills on the realities that must have clearly changed with the creation of each of these colossal structures. Their earliest mathematics and geometric skills were astounding, compared to the simplest measurement tapes and wooden devices available to them. It was a trial-by-error learning process that had been perfected to a science by the time the Old Kingdom emerged in 2686 BCE. They brought to their diverse projects an indomitable and contagious can-do attitude.

As each obelisk-raising project came to an end, the sense of group cohesiveness and national identity grew stronger. There was a wholeness and certainty of purpose, plus an increasingly overwhelming feeling of excitement as each of these edifices' pyramidion tip would seem to touch the sun above, in one more supreme endeavor and gesture to Amun-Ra. It was the penultimate climactic spiritual beginning that, by pleasing Amun-Ra, they *knew* in the deepest recesses of their being they would, in time, lead their pharaoh, nation, communities, and themselves to the salvation they strongly sought in the afterlife. That, from all evidence gathered so far, seems to be the *real* "undiscovered technology" that existed at the time. It was the technology of the earliest civilized representatives of the human race, putting its supreme individual and collective willpower to the task and creating an immortal society and civilization in an ever so secular world.

References and Further Reading

Davis, Vivien, and Renee Friedman. *Egypt Uncovered*. London: Stewart, Tabori, and Chang, 1998.

Habashi, Labib. *The Obelisks of Egypt*. London: Dent, 1978.

Iverson, Eric. *The Myth of Egypt and its Hieroglyphs in European Tradition*. Copenhagen: Gad, 1961.

Iverson, Eric. *Obelisks in Exile*, 2 vols. Copenhagen: Gad, 1968.

Lehner, Mark. *The Complete Pyramids*. London: Thames and Hudson, 1997.

Lauer, Jean Phillipe. *Histoire Monumentale des Pyramides d'Egypte*. Cairo: Institut Francais d'Archeologie Orientale, 1962.

Lauer, Jean Phillipe. *Le Mystere des Pyramides*. Paris: Presses Universitaires de France, 1988.

Lauer, Jean Phillipe. *Les Statues Ptolemaiques du Sarapieion de Memphis*. Paris: Presses Universitaires de France, 1955.

Lehner, Mark. *The Complete Pyramids*. London: Thames and Hudson, 1997.

Major, Howard. *The Domestic Architecture of the Early American Republic*. Philadelphia: Lippincott, 1926.

Menu, Bernadette. *L'obelisque de la Concorde*. Paris: 1987.

Toynbee, Joceyln M. C. *The Hadrianic School: A publisher Chapter in the History of Greek Art*. Rome: L'Erma di Bretschneider, 1967.

Trigger, Bruce G. *Nubia under the Pharaohs*. London: Thames and Hudson, 1976.

Ward-Perkins, John Bryan. *Cities of Ancient Greece and Italy: Planning in Classical Antiquity*. New York: Braziller, 1974.

Wilson, Samuel. *A Guide to the Architecture of New Orleans, 1699–1959*. New York: Reinhold, 1959.

Witt, Reginald Eldred. *Isis in the Graeco-Roman World*. London: Thames and
    Hudson, 1971.

CON

The ancient Egyptians may not have been the first civilization to produce mon-
umentally sized construction projects—the earliest civilizations of Mesopotamia
claimed that distinction—but certainly Egypt produced the greatest number and
the most awe-inspiring architecture in the preclassical world. The Egyptians
built temples to their gods, monuments to their ruling dynasties, and magnifi-
cent burial chambers for their pharaohs that continue to astound us after 5,000
years. Among the most impressive of ancient Egypt's monumental constructions
are the numerous obelisks that were erected in upper and lower Egypt. These
obelisks were often meant as expressions of divine reverence, but mostly as
commemorations for its rulers.

Obelisks from ancient Egypt fascinated the European West from the days of
the Roman Empire. Obelisks were transported from Egypt and reerected during
the classical period and again during the days of the high Renaissance. Baroque
Rome was decorated with several by its popes during the 16th and 17th centu-
ries. Later, in the 19th century, as Egyptology entered its most popular period,
obelisks were removed from Egypt and transported to the imperial capitals of
London, Paris, and New York. The massive engineering projects that were
required to transport these obelisks were well covered by the European and
American periodical press. The huge frames and pulley systems that allowed
modern engineers to erect the massive obelisks are carefully illustrated and
documented for posterity. We are also fortunate to have extensive illustrations of
the engineering methods used during the baroque period, which depict similarly
elaborate frame and pulley systems lifting the ancient stones to the amazement
of Roman crowds. There are even some relief carvings from ancient Rome that
provide at least the basic ideas of how obelisks were transported and reerected
in classical times. Unfortunately, however, there is no definitive depiction or epi-
graphic evidence (that is, ancient written evidence carved in stone) from ancient
Egypt that records how the Egyptians managed to erect and position such mas-
sive monuments. This leaves archaeologists and historians of technology with a
frustrating, but tantalizing, mystery. How did the ancient Egyptians manage to
transport and then lift these obelisks into an upright position?

The Obelisks

Obelisks were carved out of quarries, most often of red or black granite, which
can be found in the area near Aswan. They were removed from the mass of rock
by using extremely hard balls of dolerite, which were inserted into crevices and

then pounded downward until the appropriate depth was reached. The dimensions of the obelisks could be staggering in size. They were often 120 to 150 feet tall and between 12 and 14 feet across at their widest point at the bottom. It is most important to remember that these obelisks were created as one complete unit of granite, rather than being constructed out of successive blocks. The weight involved in such a huge monument was immense. The Luxor obelisk, for example, weighs some 227 tons, or nearly half a million pounds. Understanding the scale of these monuments can help us to appreciate the extraordinary difficulty involved in lifting an obelisk out of the quarry, transporting it to the desired location, and then lifting it up from its side until it was upright on just the right spot. Some have projected that the Egyptians must have possessed a technology we don't know about to repeatedly accomplish such an amazing feat. These could have been such devices as machines, pulley systems, or even kites. But this assumption does not stand up to the remaining evidence. Only one theory about the engineering techniques for erecting obelisks can be squared with everything else we *do* know about ancient Egyptian engineering and society.

The Tilted-Sand Technique and Its Proponents

The most widely accepted method that has been proposed for erecting obelisks is known as the "tilted-sand" technique. It was first formally proposed in print by Egyptologist Reginald Engelbach (1923) in his book *The Problem of the Obelisks*. Engelbach helped define the tremendous challenge the Egyptians faced and then set out to calculate how this could have been accomplished. His proposal involved both the tilted-sand design approach and the calculations for manpower required.

The obelisk, once carved out of the Aswan granite quarries, was lifted out via crowbars and "rockers" and then lifted with ropes and prying poles. Once out of the quarry, it was dragged down the slopes to the Nile. This transport was carried out most likely on wooden sleds pulled by men with ropes. There is evidence of this sled method for transporting extremely heavy cargo in the Egyptian painting and carving record. Once at the Nile, it would seem highly likely that the huge obelisk was transported down river by a massive barge. There is again evidence for this in the epigraphic record, with a famous depiction in the Temple of El-Deir el-Bahari from ancient Thebes. The carving/painting shows a massive barge with the two obelisks of Queen Hatshepsowet strapped onto its deck. Additionally Labib Habachi, in his book *The Obelisks of Egypt*, discusses one of his own discoveries of an inscription by Sabni, the governor of Aswan during the 6th dynasty. According to Habachi, Sabni recorded that he had gone south to Nubia to construct large ships specifically for the purpose of transporting obelisks. His inscription goes on to say that he was successful in his project, although Habachi tells us that no sign of those obelisks has ever been found.

Once the obelisk arrived at its intended site, workers constructed a large earthen ramp, which ended, at its highest point, immediately in front of the intended location for the upended obelisk. The earthen ramp was supported by brick and timber retaining walls and filled with soil and sand. The obelisk then was dragged up the ramp, bottom first and most likely without the aid of a sled, until it reached the end of the ramp and its bottom section overhung the ledge. At that point laborers grasped hundreds or thousands of ropes and pulled from the opposite (top) side as a stabilizing force, as other laborers continued to pull the obelisk slowly over the edge. The greater weight of the bottom of the obelisk generated enough gravity, as it "fell" over the side of the sand rampart, to elevate the top of the obelisk, flipping it slowly upright as it descended. The pit into which the obelisk was sinking and in which it would eventually rest on its base, was filled with sand. Other laborers then cleared out the sand systematically, with the obelisk able to continue its slow, regulated descent as the sand disappeared.

The bottom of the obelisk slowly descended the side of the rampart and into the sand-filled pit, and when it reached the bottom, the inside edge of the base was fit into a groove carved in the base stone. With the obelisk's bottom edge securely in that groove, sand then was cleared out from under the base, allowing the obelisk to rotate downward until it rested with its base flush against the base stone. The pit was then cleared of all excess sand and the ramparts taken apart and removed, leaving a freestanding obelisk in an open space, or very often in a temple. This would have required massive amounts of manpower to accomplish, but it would not have required more complex engineering mechanisms, like pulleys. Engelbach, for example, calculated that to pull the Luxor obelisk up a modest ramp would have required the force of 2,000 men.

Since Engelbach, other scholars have endorsed his basic idea while adding small refinements. Dieter Arnold (1991), for instance, believes that the scale of the ramp projected by Engelbach, and later Henri Chevrier, was too great to have been workable in already-existing temples. Engelbach and Chevrier projected that an obelisk would slide down the gradual angle of a sand ramp that was nearly as high as the obelisk itself. Arnold proposes that the ramps could have been much smaller than this, only perhaps a third the height of the obelisk, and the obelisk could have pivoted on the edge and into the pit, rather than needing to slide down the side.

The Problem of Evidence

Regardless of the finer points of execution, the general idea of the tilted-sand technique is the most widely proposed and accepted theory by scholars. The principal reason for this is that the tilted-sand method is consistent with other areas of Egyptian life. Still, it must be emphasized that no direct evidence for

An unfinished obelisk at Aswan, Egypt. (iStockphoto)

this method exists. There are mountains of recorded inscriptions, temple drawings, and paintings from ancient Egypt, but no clear depiction of the process for erecting of obelisks has been found. There is, however, one scene that deals with the subject and illustrates the "symbolic erection" of two obelisks, included in Arnold's book. This is the pictorial inscription of Ptolemaios XII pulling two obelisks, which are at a tilting angle, upright. Around the obelisks are two pieces of rope, and Ptolemaios XII is pulling one rope in each of his hands. The obelisks are sitting atop a flat base stone, but no indication of how they pivoted into place is included in the rather plain illustration. Though not definitive, this representation does incline toward the method of human labor pulling obelisks into place with rope. Of course no pulleys or frame constructions are depicted, but neither are ramps nor sand pits. Absolute proof of the method for erecting obelisks will have to await the discovery of new epigraphic or archaeological evidence. But the tilted-sand theory can be tested somewhat by examining it within other ancient Egyptian contexts. Is the tilted-sand theory consistent with what we know about Egyptian construction methods at the time of the earliest obelisks? Did other contemporary civilizations in contact with Egypt use more complex technologies for such lifting in their own monumental building? Finally, are the enormous labor requirements consistent with what is known about Egyptian societal norms and the cultural meaning assigned to the obelisks?

The Context of Egyptian Building Methods

The first of these contexts is that of the extant evidence of Egyptian construction and architectural technology. It is in assessing this that Dieter Arnold's work *Building in Egypt* is exceptionally important. Arnold brings together most of the important work and evidence surrounding this topic and includes his own projections. Some of these are solid endorsements of the tilted-sand method. The most important question is that of ramparts. Did the ancient Egyptians use ramps, such as those suggested by Engelbach and Chevrier, in their other monumental building projects? This can be answered with a definitive "yes." Somers, Clarke, and Engelbach (1930) endorsed this method for the building of pyramids in their book *Ancient Egyptian Masonry*. Arnold lists several examples of ancient Egyptian ramp building and, significantly, includes photographs from the archaeological sites. Most ramps, of course, would have been removed when the work was done, and so this is a difficult area to discern, but some remains do still exist. Arnold includes the ramp at the pyramid of Amenemhat I at Lisht and the ramp at the pyramid of Senwosret I among his examples of excavated ramparts. He explains that these ramps were made mostly of bricks and Nile mud, though they were held together by brick retaining walls. They were quite long to allow inclining at fairly gradual angles, an obvious advantage when pulling a massive stone uphill. Ramp building then was a common practice in the construction of monumental architecture and would present no aberration in being part of the process for erecting obelisks.

The tilted-sand method is also consistent with our knowledge of Egyptian methods for lifting heavy stones. There were three principal methods for this. First, the lugging of heavy cargo up an inclined plane such as the ramp being pulled by ropes directly attached to the cargo or sleds. Next, Egyptians used ropes to lift and may well have used wooden frames to assist in this process. There have been several instances of hollow holes in excavated ramps and construction areas that would accommodate wooden posts. What exactly these were used for is still in question, and it brings up the tricky question of pulleys in ancient Egypt.

For many decades Egypt was assumed to have had no knowledge of pulley systems, but this is becoming less clear. E. A. Wallis Budge (1926), a giant in the field of Egyptology, in his book *Cleopatra's Needles and Other Egyptian Obelisks*, rejected the method of pulleys in moving and erecting obelisks. There was simply no evidence for this in the archaeological record, according to Budge. J. J. Coulton, a renowned historian of technology, also rejected any evidence of pulleys in Egypt. Arnold, however, in his more recent study, includes examples in his work of quite primitive wheel pulleys, the oldest dating from the end of the 12th dynasty, or approximately 1800–1700 BCE. These primitive wheels, though, are not seen by Arnold as part of genuinely load-bearing pulleys, but instead merely as instruments for changing the direction of pull. Given

these facts, it would seem likely that the posts along construction ramps were part of wooden frames that accommodated proto-pulleys that allowed laborers to pull from the ground and lift a stone partially upward, while others pulled from the front to advance it more easily. This is, however, conjecture without any written or pictorial evidence. But, despite these primitive rope wheels or proto-pulleys, it remains the most widely acknowledged view that the Egyptians did not have the advantage of weight-relieving pulley systems to aid them in the erection of obelisks.

The third acknowledged method the Egyptians used for lifting heavy loads was levers. This method was almost certainly used to lift obelisks out of their quarry. For use in the tilted-sand technique, levers plausibly could have been used to tip the obelisk as it approached the edge of the ramp and was ready to pivot. Regardless, the tilted-sand technique would certainly not appear to be inconsistent with the ancient Egyptians use of ramps, the use of ropes and simple frames or wheels, or the use of levers.

Another method of moving heavy loads in the ancient and medieval worlds was the employment of animal power. However, this has been discounted by scholars in the case of the ancient Egyptians and their monumental building. One possibility that comes to mind, given the geographic location of Egypt, is the employment of elephants in pulling enormous loads. Elephants, however, were no longer indigenous to Egypt after approximately 2900 BCE and were never common enough, even if imported, to have made a contribution to construction techniques. Egyptologist Mark Lehner asserts that there is evidence for the use of cattle in labor projects, but the numbers that would have been required to move obelisks would have been very difficult to manage. Both logic and pictorial evidence of monumental construction suggests that it was human labor that was more abundant and easier to command.

Finally, there is another important question bearing on the ancient Egyptians' ability to have used more complex technologies—their use of metals. To allow a pulley system to be effective, given the extraordinary weights of the obelisks, the axles for the pulleys would have had to be very strong. To attain that level of strength, Egyptians would have needed iron or steel. The Egyptians, however, erected many of their most prominent obelisks before the Iron Age ever reached Egypt in the first millennia BCE. The strongest metal in general use before that time was bronze, and this would not have produced pulley system components of sufficient strength to support the kind of weight involved in transporting or lifting obelisks.

The Context of Contemporary Civilizations

Another way of establishing connections to Egyptian building methods is to examine the methods of contemporary civilizations that were in contact with

---

The Shedu of Assyria

The benevolent demons of Assyrian mythology most often portrayed in art are the *shedu* and the *lammasu*—winged bulls and winged lions, both with human heads. These liminal creatures were the protective spirits of dwellings and cities, and as such were carved in bas-relief on a variety of available surfaces. They always appeared in pairs—generally one at each side of an entrance—and small representations of *shedu* could be buried beneath the thresholds of private homes to ensure the safety and well-being of those who lived there. They were also the subjects of statues, sometimes enormous statues. Famous examples are kept at the Louvre, Chicago's Oriental Institute, and the Metropolitan Museum of Art.

Egypt. One important example is the Assyrian Empire. Scholar Robert Heizer (1966) assembled a comparative study of ancient cultures and their engineering methods in his article "Ancient Heavy Transport, Methods and Achievements." Heizer opined that the Egyptians did not possess technology beyond the lever system and sled transports for heavy architecture on land. He also found pictorial evidence in Assyrian inscriptions of the Assyrian methods for transporting great monuments. A panel excavated from Nineveh depicts mass labor moving one of the famous Assyrian winged bulls by hundreds or thousands pulling on ropes directly attached to a great sled. Other laborers hold ropes attached to the winged bull itself, which sits on the sled, presumably to provide stabilization and prevent the great sculpture from tipping. Behind the sled, laborers use a basic wooden lever to lift the back to make pulling the sled forward less difficult. Finally, long lines of other laborers and carts of extra poles and rope follow behind. It is a remarkable image that confirms the Assyrians of the 8th century BCE, many centuries *after* the Egyptians began to move obelisks, routinely organized their projects using mass labor forces and simple ropes, sleds, and levers.

During the same period, similar techniques were used in archaic Greece. Scholar J. J. Coulton (1974) studied ancient Greek building and lifting methods. In his article "Lifting in Early Greek Architecture," Coulton identified great similarities between the two cultures, including some that have influenced our ideas about the limitations of the ancient Egyptian ability to lift massive obelisks.

Coulton reviewed such ingenious developments of the early Greeks as the use of the U-shaped hole to lift. This method consisted of drilling a U-shaped hole into a heavy block of stone, threading a rope through it, and then attaching the two ends of the rope to a pole. That pole was then used to lift the block by placing it on another log and pressing on its far end, creating a lever mechanism. Despite this kind of advance, Coulton found early Greek methods to be generally primitive and quite similar to Egyptian methods of the time. He finds no evidence for weight-relieving pulley systems in either early Greece or Egypt. He was

convinced the heaviest lifting was done by wooden levers and not cranes or hoists. He also found some similarities in the Greek use of temporary ramps, which, he asserted, was the most widely used method of the Egyptians. He concluded that the early Greeks possessed no winch or pulley systems before the sixth century BCE and that their methods were quite similar to the Egyptians', from whom they had learned the techniques of ramps and levers. The appearance of complex pulley technology, then, according to Coulton, was a later Greek advance and cannot be credited to the older Egyptian cultures, which Coulton maintains never developed such tools. So neither the Assyrians nor the Greeks were using technology beyond simple ropes, sleds, ramps, and levers during the 8th century BCE, even though both civilizations were certainly in contact with the Egyptians and had presumably derived their technologies from Egypt over the centuries.

The Context of Egyptian Society

Another way of evaluating the plausibility of the tilted-sand technique is to examine whether the organization of labor for such a project is consistent with the values of ancient Egyptian society. In assessing the force required for moving a 224-ton obelisk up a temporary construction ramp, Engelbach estimated that 94 tons of force would be necessary to move it in short, repeated intervals. To generate this level of human force would have taken the efforts of some 2,000 laborers pulling and pushing the obelisk. Is there evidence in the written or archaeological record to support these kinds of mass-labor projects?

It has become axiomatic that the ancient Egyptians used thousands of forced laborers, probably slaves, to perform the backbreaking work of constructing monumental structures such as the pyramids of Giza. Newer work on the plains of Giza by Egyptian director of Giza excavations Zahi Hawass and Lehner is changing our view of these projects. The conditions for laborers appear to have been far better than had been supposed; these laborers appear to have had their families living with them and to have received medical care. Therefore, the view that these were slave laborers has come under dispute. A different view is emerging that these may have been willing volunteers working in a prestigious assignment to the glory of gods and the pharaohs. This new evidence may point to a religiously motivated labor force similar to those communities that built the medieval cathedrals in western Europe.

Whether willing communities of laborers undertook these projects or not, one thing the excavations at Giza do support clearly is that massive numbers were involved. Hawass projects that 36,000 laborers may have been working on the great pyramids at any one time, and that laborers were constantly being rotated in and out of the workforce. There is ample pictorial/epigraphic evidence for mass labor projects as well. Some who contest the tilted-sand method

and project that the Egyptians possessed more refined technology do so based on the assumption that the Egyptians would have sought more efficiency in their building. However, unambiguous depictions exist of enormous stone monuments being transported and moved into place by mass labor, using the simple rope and sled method. One of the most famous of these depictions is a 12th dynasty tomb painting at El Bersheh dating from approximately 1800 BCE. The painting shows a massive alabaster statue of Djehutihetep being pulled on a sled by a mass of laborers pulling on ropes directly attached to the sled. Atop the statue a work leader claps cadence for the laborers. There is no sign of pulleys or any technologies that would relieve the need for thousands of laborers, but instead a massive workforce is depicted in organized columns with a corps of others carrying extra timber and water.

The use of mass labor is apparent in Egyptian pictorial records, although no examples of labor-saving technologies, beyond those already discussed, have come to light. This use of mass labor would also be consistent with the meaning and intent of obelisks. Obelisks were dedicated almost exclusively to Egyptian gods and placed in temples or to pharaohs to commemorate the accomplishments of their rule. These monuments then were built and erected as permanent memorials, according to William A. Ward (1966) in his article "They Built for Eternity in Ancient Egypt." Ward cites written evidence from the records of Ramses II that temples were explicitly meant to be places for gods to occupy forever. The use of massive labor forces then would have certainly been worth the effort and human costs in the Egyptian worldview, according to Ward. With the pharaohs' virtually absolute power over the people and in light of the religious meaning of these constructions, the inefficiencies of mass labor required in lifting obelisks would have been a minor issue by comparison.

Tilted Sand in Practice and Experimental Archaeology

Finally, to establish that the tilted-sand technique is the most plausible theory and that this method can work in practice should be proved. Methods proposed in scholarly articles and diagrams are useful and necessary, but the final determination as to whether this method was practicable can only be made through experimental archaeology, the practice of duplicating proposed functions or manufacture of material artifacts while re-creating ancient conditions as closely as possible. This process often results in a new appreciation of the difficulties our ancient ancestors faced and the ingenuity they displayed in meeting these challenges. Fortunately some archaeologists have undertaken to test the tilted-sand method.

Lehner and fellow archaeologist/engineer Martin Isler each took different approaches to using the tilted-sand method in an experiment in 1994 and again in 1999. Lehner and his group were part of a project produced by the television

program *Nova* in which they attempted to transport and pivot a 40-ton obelisk. In 1994 that team was successful in moving the obelisk, with the help of many local experienced Egyptian workers, into position and pivoting it onto a groove in its side until it rested on the side of the sand hill. However, Lehner's team was unsuccessful in pulling the obelisk the rest of the way over into an upright position. With time, it would appear this would have been accomplished, but the television show's production schedule did not allow the team to continue work. Eventually, however, another team's subsequent effort was successful. This proves that the method is practicable, although extremely difficult. It should be said that the obelisks used by the various experimental archaeologists were far smaller and lighter than the largest erected by the ancient Egyptians, which could be as heavy as 400 tons. Thus the experiments conducted by Lehner and Isler can be seen as quite convincing in terms of the engineering and techniques, but they cannot be decisive until and unless obelisks of similar size and similar-sized work crews are used. Given the restrictions of resources from university funding or television programs and the limits of space and human resources, an experiment on that scale is unlikely to be undertaken.

Other Propositions

There are some remarkable new projections as to how the ancient Egyptians erected the obelisks. The most imaginative of these is the notion that the ancients could have harnessed the power of the wind through the use of a kind of kite. A group at the California Institute of Technology has worked on this theory for nearly five years and has run some successful experiments and simulations with smaller obelisks. A great kite made of solid material would be connected by ropes to a "clutch and frame" brake mechanism on the ground in front of the obelisk. The main rope from the kite would pass through the brake and connect to the top of a tall wooden frame that housed a series of weight-bearing pulleys. Passing through the pulleys, the ropes would be connected directly to the top of the obelisk. As the wind filled the kite, it would move, pulling the ropes and eventually lifting the head of the obelisk until it could be positioned completely upright.

The problem with such possibilities, however, is that they do not appear to be consistent with existing evidence of Egyptian society. Pulleys mounted on frames, all of immense strength, would be necessary to transfer the power of the wind to the obelisk itself. As we have seen, no convincing evidence exists to suggest the Egyptians had mastered pulleys during the preclassical age. Further, no evidence exists to suggest that the ancient Egyptians were familiar with aerial technologies like kites or balloons. So while such theories are fascinating and can even be proved workable, they are not consistent with the evidence we have about Egyptian society, materials, labor, and construction methods.

Conclusion

It is a tantalizing and frustrating engineering riddle how the ancient Egyptians, as early as the Old Kingdom, raised stone obelisks of up to 400 tons. No conclusive textual or epigraphic evidence exists to tell us beyond a doubt how the ancient engineers solved such an immense problem. In the absence of such evidence, we can only attempt to construct the most plausible way this could have been done. But, to be truly plausible, any theories about such engineering methods must be consistent with the conclusive evidence we do possess about other areas of Egyptian life.

Engelbach's proposal of the tilted-sand method provided a method that appealed to our common sense in functional terms. Tipping the obelisk over a sand ramp until it rested over its intended location and then gradually pulling it into place (or removing sand from a pit) to allow a gradual descent appears logical and workable on a diagram. But, it is when we begin to analyze this method within the context of other fragments of evidence that it becomes such a persuasive candidate. The tilted-sand method would have required massive amounts of human labor as an energy source. This is perfectly consistent with the evidence that exists from other large construction projects in the Egyptian world. Textual and pictorial evidence exists that indicates that enormous crews of laborers were typical for large projects. Archaeological excavations, such as those at the labor camps of Giza, support this beyond question. Conservation of human resources does not appear to have been a driving motivation when projects were undertaken to honor pharaohs or the gods. We also have pictorial and archaeological evidence of the equipment used by such laborers. Their methods appear to quite simple, using ropes directly attached to heavy objects or transport sleds with human laborers pulling at the ends. No evidence of pulleys or more elaborate technologies is available, but there is ample evidence of the use of ropes, sleds, earthen ramps, and levers. Further, the Egyptians lacked materials like iron and steel to have made pulleys genuinely productive. Additionally, no evidence exists to suggest that the Egyptians used draught animals consistently as a source of power. Elephants were not common enough in Egypt to have made a meaningful contribution, and there are no depictions of cattle or oxen directly involved in monumental architecture.

Finally, we can see the values of Egyptian society embedded within the methods involved in the tilted-sand technique. Obelisks were constructed and erected as permanent memorials to pharaohs and gods and to last through eternity. Given the divine status of pharaohs and their absolute power over the populace, they had no misgivings about the tremendous human and nonhuman resources involved in these projects. The kind of mass labor necessary for the simple technologies of the tilted-sand method would have been routine in pharaonic construction, while the motivation to develop ingenious systems of pulleys or even kites would have been highly exceptional.

References and Further Reading

Arnold, Dieter. *Building in Egypt: Pharaonic Stone Masonry*. New York: Oxford University Press, 1991.

Budge, E. A. Wallis. *Cleopatra's Needles and Other Egyptian Obelisks*. London: Religious Tract Society, 1926.

Clarke, Somers, and R. Engelbach. *Ancient Egyptian Masonry: The Building Craft*. London: Oxford University Press, 1930.

Coulton, J. J. "Lifting in Early Greek Architecture." *Journal of Hellenic Studies* 94 (1974): 1–19.

Engelbach, R. *The Problem of the Obelisks*. New York: George Doran, 1923.

Heizer, Robert. "Ancient Heavy Transport, Methods and Achievements." *Science* 153, 3738 (1966): 821–30.

Habachi, Labib. *The Obelisks of Egypt: Skyscrapers of the Past*. New York: Charles Scribner's Sons, 1977.

Isler, Martin. *Sticks, Stones, & Shadows: Building the Egyptian Pyramids*. Norman: University of Oklahoma Press, 2001.

Ward, William, A. "They Built for Eternity in Ancient Egypt." *Saudi Aramco World* 17, 2 (1966): 1–6.

12

The Beta Israel (or Falasha) People of Ethiopia are one of the Lost Tribes of Israel.

PRO   Barry Stiefel
CON   Talaat Shehata

PRO

The view that the Beta Israel people of Ethiopia, often called the Falasha, is one of the 10 Lost Tribes of Israel is a notion asserted by the Beta Israel, the Rabbinate of Israel and the Israeli government, and by various other groups of African and Jewish affiliation. Others argue a different origin for the Beta Israel, either from a migration of Jews long after the biblical period or through conversion of local Ethiopian peoples. Based on the available historical record, the evidence does suggest that the Jews who settled in Ethiopia came from multiple sources. Examples include origins in southern Arabia and the upper regions of the Nile River in Egypt, local Ethiopians who converted and become part of the Jewish population, Ethiopian Christian heretics who chose an expression of Mosaic monotheism, as well as others based on local myths and legends (Kaplan 1992: 209–11; Quirin 1993: 301–2; Pankhurst 1992: 567). However, the historical record on the arrival of Judaism in Ethiopia is unable to exclusively point to the specific instance in history for the establishment of this first settlement. Unfortunately, the known historical record dwindles down to a trickle the farther one goes back into history. By means of methodological deduction, the question begs to be asked: When did the Jews first come to Ethiopia? The answer to this question may resolve the mystery of whether the origins of the Beta Israel could have come from one of the 10 Lost Tribes of Israel.

The purpose of this section is not to refute the historical evidence that claims that the Beta Israel descended from postbiblical-period Jewish migrations from Egypt or Arabia, or if some of the local inhabitants converted to Judaism. The argument is only that there are reasons to suggest that the first Jews to settle in Ethiopia may have done so during the biblical period, at a time when the Israelites still identified themselves as 12 different tribes within a nation. Research has revealed that the origins for the Beta Israel are very ancient, possibly so old that the initial Jewish root of it may date from the biblical period, and this creates a possible connection of descent with one of the ancient 10 Lost Tribes of Israel. In respect to these other explanations, from Egypt or Arabia, or through

271

conversion of members of the local population, there appears to be sufficient historical evidence to support that these events may have happened as studied by other historians. In fact, these findings do much to explain the origins of the cultural and religious practices of the Beta Israel in the present and why they differ in the ways they do from more familiar mainstream Judaism—as an infusion or addition to an already present Jewish root population and culture.

Furthermore, it should also be clarified that the 10 Lost Tribes include Asher, Dan, Gad, Issachar, Naphtali, Rueben, Simeon, Zebulon, and the half-tribes of Ephraim and Manasseh (which together make the whole Tribe of Joseph). The "un-lost" tribes are Judah, Benjamin, and Levi. During the biblical period, Ethiopia was also often known as the land of Cush.

Beta Israel Cultural Traits and Religious Traditions

The most important work within Judaism is the Torah, which is often called the Bible or Old Testament in non-Jewish circles. Among the Beta Israel this work is called the Orit, which means the same in the Ge'ez language—the language historically spoken by many Beta Israel. In mainstream rabbinical Judaism the biblical canon is called the Tanakh, which includes the Torah (the five books of Moses—Genesis, Exodus, Leviticus, Numbers, and Deuteronomy)—as well as the prophets (the books of Joshua, Judges, Samuel, Kings, Isaiah, Jeremiah,

Ethiopian Jews pray in Jerusalem during the Sigd holiday in November. The prayer is performed by Ethiopian Jews every year to celebrate their community's connection and commitment to Israel. (AP Photo/Emilio Morenatti)

Ezekiel, and the minor prophets) and writings (the books of Psalms, Proverbs, Job, Songs of Songs, Ruth, Lamentations, Ecclesiastes, Esther, Daniel, Ezra, Nehemiah, and Chronicles). The Beta Israel have all of these books as part of their biblical canon in addition to the apocrypha books of Enoch, Jubilees, and Baruch. Furthermore, the Beta Israel have their own unique texts that they study on the lives of the prophets Adam, Abraham, Isaac, Jacob, Moses, Aaron, and Elijah; as well as on topics pertaining to the sabbath, priestly responsibilities, the calendar, and holidays. Important holidays in the Beta Israel include Rosh Hashanah (New Year Festival), Yom Kippur (Day of Atonement), Succoth (Festival of Booths), and Passover (Festival of Matzoth); however, they are all known by their respective Ge'ez names. Historically, these are the oldest holidays within Judaism and they are shared by mainstream rabbinical Judaism. However, the Beta Israel lack more recent holidays, such as Hanukkah (Festival of Lights), which dates from the second century BCE. Furthermore, the Beta Israel have a holiday unique to themselves called Sigd, which occurs in the month of November (specifically the 29th of the month of Cheshvan in the Hebrew calendar), which celebrates the Orit. The Beta Israel also follow the practice of *kashrut*, which has more similarities with the food taboos described in the Bible (mostly defined in Leviticus), compared to the more intricate *kashrut* practices of rabbinical Judaism, which are expanded upon in the Talmud. Finally, the Beta Israel also observe rules of ritual purity in a similar fashion to mainstream (orthodox) Judaism, specifically in regard to a married women's menstruation cycle and the ritual immersion in a bath at the end of the cycle before returning to her husband, as well as the custom of male infant circumcision (Weil 1991: 30–40).

Judaism of the Beta Israel Compared to Mainstream Rabbinic Judaism

Religion for the Beta Israel, before the 20th century, hardly resembled mainstream Judaism as it is known today. Beginning in the 20th century this began to change as the isolation of the Beta Israel became diminished with frequent contact with other Jews, especially with the Israeli airlifts of Ethiopian Jews, called Operation Moses and Operation Solomon, which took place in 1984 and 1991, respectively. Historically, Beta Israel cultural and religious practices of Judaism were very different from the stereotypical image of the Eastern European orthodox Hassidic man wearing a black suit, a tall fur hat, with a long beard and long curly *payes* (sideburns); or even the young American reform urban professional female, with a colorful prayer shawl and ornate yarmulke. Though worlds apart in their thought and philosophy on what Judaism is—between ultra-orthodoxy and liberal reform—both of these expressions of Judaism fall within the larger umbrella of rabbinical Judaism based on the precepts of the written Bible (called the Tanakh) and the oral Bible (which was later written down and called the Talmud).

Though almost monolithic today, Judaism was once more diverse in its expression, with historical examples including the Sadducees, Pharisee, and Essenes; with the one surviving example being the Karaites. The fundamental difference between all of these Jewish groups versus rabbinical Judaism (which includes the orthodox, conservative, reform, and the reconstructionist movements) is the use of the Talmud and the institution of the rabbi as the religious and spiritual leadership of the people. Many of these different expressions developed (and fell) during the last centuries BCE and the early centuries CE. With the fall of the Roman Empire during the 400s CE, rabbinical Judaism had become entrenched as the primary expression of Judaism for the majority who practiced the religion. This rise to prominence filled the void left by the destruction of the second temple in 70 CE, during the Judean revolt against the Roman Empire. With the destruction of the second temple came the end of an era in Judaism that had been focused on the sacrificial practices centered in the second temple and its predecessor the first temple (often called King Solomon's temple). The temple of King Solomon was destroyed by the Babylonians in 586 BCE. Rabbinic Judaism replaced the practice of animal (livestock) sacrifice with verbally spoken prayers performed in a new institution called the synagogue. The synagogue and prayers served as a substitute for the sacrificial practices of the temple, and so rabbinic Judaism continues to use prayers as the means of worship and the synagogue as the place worship (Barnavi and Eliav-Feldon 1992: 30–37, 42–43).

Presuming a settlement date toward the end of the biblical period for the ancestors of the Beta Israel in Ethiopia, this arrival would have predated the rise of rabbinic Judaism as the preeminent practice of the religion. Due to the remoteness of the region in Ethiopia, where the Beta Israel lived during the historical period, rabbinical Judaism was not in a position to exert much authority or influence over the Beta Israel as to how they should conduct religious practices following the destruction of the second temple. However, this is not to say that the Judaism practiced by the Beta Israel is a surviving relic of the original sacrificial cult of 2,000 years ago. It is known that during the historical period, cultural influences from the Beta Israel's neighbors, who were Christian, Muslim, and various pagan groups, also had some impact. One of the most significant examples recorded took place during the 15th century with Abba Sabra and Abba Amlak, two Christian monks who fell out favor with the Christian Ethiopian Orthodox Church and converted to Beta Israel Judaism. These monks infused some Christian Ethiopian orthodox-like interpretations of the Orit after their conversion into Beta Israel Judaism because of their Christian background and prior education from the Old Testament (Quirin 1979: 239–42). Therefore, the Judaism as practiced by the Beta Israel has to be seen within the context of its near isolation from mainstream rabbinic Judaism, impacts from its non-Jewish neighbors, and its own independent development over the millennia. This is also why the

Beta Israel lacked understanding of the Hebrew language in the 20th century, before contact increased with the outside world. Since there has been more frequent contact between the Beta Israel and mainstream Judaism during the 20th century, Beta Israel Judaism has changed to reflect rabbinical influences.

Available Historical Record

Reconstructing the historic past of Judaism in Ethiopia is wrought with difficulty because of the scarcity of written record, which forces the researcher to rely on outside sources that are sometimes controversial. The earliest known available records date from the fourth century from early Christian missionary accounts. However, according to these earliest Ethiopian records and traditions, up to half of Ethiopia was Jewish at the time that Christianity spread into the region (Kessler 1982: 3, 77). The early modern historical record on the Beta Israel within Ethiopia begins in the 14th century, almost 1,000 years later, where the Beta Israel are mentioned as having participated in local military conflicts. During the 1410s, the Christian Emperor Yeshaq declared war on the Beta Israel, and following his victory over them he declared: "He who is baptized in the Christian religion may inherit the land of his father otherwise let him be a *Falasi* [a landless person, a wanderer]" (Kaplan 1992: 216). This is the first recorded mention of Falasi directed toward the Jews in Ethiopia, the root word for Falasha, which refers to someone without land or a foreigner. The term Falasha often has pejorative connotations in Ethiopian society.

Up until 1624, with the defeat of the Beta Israel by the Christian Emperor Susenyos, the Beta Israel had some level of autonomy. According to legend, the Beta Israel historically ruled an expansive amount of territory within what is

---

Prester John and Ethiopian Christianity

During medieval times in Europe, Christians told legends of Prester John, often believed to be the emperor of Ethiopia, and shown on maps of the later Middle Ages and the early modern period. Many hoped that he would be an ally of Christian Europe against the Turks during the Crusades, and later against the Ottoman Empire.

The story of Prester John rests on the traditions surrounding Thomas the apostle, who was said to have ventured around that area and converted the people there to Christianity. His success was said, in part, to stem from a Jewish population there, which was connected with the legend of the queen of Sheba. Indeed when two Portuguese explorers Pero da Covilha and Afonso de Paiva visited the region in the late 1480s, they did find a Christian civilization in Ethiopia, and many centuries later historians were able to find the Falashas, which was seen as confirmation of some aspects of the legend.

now Ethiopia. With the loss of their lands and autonomy, employment opportunities available to the Beta Israel changed to craftsman and laborers in Ethiopia's economy. This was considered less preferable to the more esteemed agricultural-related professions in Ethiopian society, which is dependent on land ownership. Thus, over the centuries, the position of the Beta Israel within greater Ethiopian society gradually deteriorated (Quirin 1979: 241–43). By the 16th century, the term Falasha in reference to the Jews of Ethiopia is found outside of Ethiopia, which appeared in a letter written in Jerusalem by Abraham ben Eliezer Ha-Levi in 1528 (Kaplan 1992: 216). What is interesting about this is that Abraham ben Eliezer Ha-Levi was also Jewish.

Prior to the 15th century, the Beta Israel were referred to as Ayhud (Quirin 1993: 299–301), a term used to refer to Jews throughout the region in multiple languages and having a Hebraic-Arabic origin (of *ya-hud*). Also during the 16th century, the chief rabbi of Egypt, David ben Solomon ibn Abi Zimra (also known as the *Radbaz*), had ruled within Jewish law that the Beta Israel were undeniably Jewish. His findings were in agreement with other rabbinical authorities at his time and before, some of which mentioned the presence of Jews in Ethiopia. The most significant example was by Rabbi Moses Maimonides (also known as the *Rambam*), who declared during the 1100s that all Jews were of the same people regardless of skin color, even though it is not known if he was aware of the Beta Israel in Ethiopia (Corinaldi 1998: 195–97).

One of the oldest surviving testimonies of a Jewish presence in Ethiopia is the account of a Jew named Eldad the Danite. Originally transcribed in what is now Tunisia during the ninth century, Eldad stated that he was the lone survivor of a shipwreck and that he originally came from Ethiopia where his people were pastoralists. He is described as having very dark skin and knowledge of a "strange sort of Hebrew," as well as a descendant of the Israelite tribe of Dan—one of the 10 Lost Tribes! Eldad's account of a Jewish presence is not found among the Beta Israel (since it was recorded outside of Ethiopia); however, there are local legends and an oral tradition that corroborate a Beta Israel origin from the tribe of Dan. Furthermore, Eldad's description also lists the whereabouts of three other "Lost Tribes"—Naphtali, Gad, and Asher—as residing in close proximity to his tribe (Adler 1987: 4–21). Eldad's description of Jewish life in Ethiopia mentions their lack of knowledge of the Talmud. However, as fantastic as the narrative by Eldad is, there are sections that must be taken with a grain of salt due to embellishment, exaggeration, or improbability. One example is his description of a river comprised of grinding stones (instead of water), which ceased its flow on Saturdays (the Jewish sabbath). These issues raise the concern of how much of Eldad's story can be accepted as fact for the purposes of historiography. However, many other noteworthy accounts from the European medieval period contain significant embellishment and fantasy, such as Marco Polo's account of various kinds of monsters in Asia in his book, *Travels*

(Jackson 1998: 82–83). Therefore, historical information can be obtained and inferred from Eldad's account, but cannot always be taken at face value.

According to Eldad and other legends of the Beta Israel, the tribe of Dan settled in what is now Ethiopia during the civil war that erupted in King Solomon's kingdom of Israel shortly following his death (around 928 BCE). This civil war concerned the succession of his son, Rehoboam, to the throne, which was objected to by many of the tribes under the leadership of Jeroboam. The end result was the splitting of King Solomon's realm into two kingdoms: Israel in the north and Judah in the south (Judah is named after the tribe that dominated this kingdom). Eldad's testimony mentions that much of his tribe left at this time for Africa. With the conquest of the kingdom of Israel in 722 BCE by the Assyrians under King Sennacherib, the 10 tribes that resided there were expelled (thus becoming the 10 Lost Tribes). It is believed that many were taken back to Assyria into captivity. However, based on Eldad's account, a significant contingent of the tribes of Naphtali, Gad, and Asher joined the tribe of Dan in northeastern Africa (Adler 1987: 4–21).

Predating the account of Eldad by about 500 years are the records of the early church historian Philostorgios, from circa 327 CE. In his account of the conversion of the peoples of Ethiopia to Christianity, he mentions the firm opposition to the missionaries of the Jews who were already living there. This date predates the short-lived Jewish kingdom of Himyar, which was located on the opposite side of the Red Sea during the sixth century CE. Upon the fall of Himyar in 525 CE (Lecker 1995: 636), many Jewish refugees are believed to have fled south into Ethiopia, bringing an infusion of the Jewish population. Therefore, there is historical evidence that supports a Jewish presence in the region of Ethiopia that predates the emergence of Islam and the spread of Christianity.

Unfortunately, the attention given to a Jewish presence in Ethiopia in the Talmud, which was written between third century CE (beginning with the compilation of the Mishnah) and seventh century CE (ending with completion of the Gemora), is relatively scant. However, considering that the Beta Israel were historically without the Talmud and are not known to have participated in the Responsa during this period or later, this is not surprising. The Responsa, as it is known today, first developed during the third century CE and is a rabbinic exegesis literature on the Tanakh, Talmud, and other aspects of Jewish religious law. In many instances, Responsa were correspondences (or responses) between communities in the Jewish Diaspora about issues that needed resolutions from prestigious rabbinic authorities, some dating as late as the 18th century. With the destruction of the second temple in 70 CE and the defeat of the Bar Kokhba revolt (132–135 CE), world Jewry found itself without a centralized node to connect and keep the diaspora communities together. Eventually, rabbinic Judaism, with its emphasis of reliance on the Torah, Talmud, and use of the synagogue, arose to fill in this void (Strack and Stemberger 1996: 1–7, 216–18). However,

it is obvious that the Beta Israel residing in Ethiopia were unable to connect with the recentralizing influences of early rabbinic Judaism and so remained isolated from the rest of world Jewry.

However, predating the first centuries CE was a very old Jewish presence in northeastern Africa that has been documented. From a very early date a sizable community of Jews is known to have resided in Elephantine, which was a prominent city in the ancient world located near the present location of Aswan, just north of the Egyptian-Sudanese border. The archaeological record for this Jewish presence dates to the sixth century BCE, which is significant because it places a substantiated Jewish presence in the region in the midst of the first expulsion by the Babylonians, when tribal identification and distinction among the Jews were still very important. In fact, the Jewish community of Elephantine was so significant in the ancient world that a separate temple from the main one in Jerusalem was built, which was used until the early fifth century BCE. During this period, Elephantine served as a military and commercial trade center on the southern frontier of the ancient Egyptian kingdoms, and many of the Jews who lived there were known to have served as mercenaries and merchants (Rosenberg 2004: 4–13). Seafaring trade routes are also known to have existed for centuries between the Middle East and East Africa that utilized the Red Sea and other coastal waters on the Indian Ocean, which could have also exposed the Jews of Elephantine to Ethiopia and its economy.

Last, but not least, within the Jewish biblical canon a Jewish presence in Cush (the biblical term for Ethiopia) is mentioned. According to the prophet Isaiah, Jews are described as living in Cush:

> And it shall come to pass in that day, that the Lord will set his hand again the second time to recover the remnant of His people, that shall remain from Assyria and from Egypt, and from Pathros, and from Cush, and from Elam, and from Shinar, and from Hamath, and from the islands of the sea. And He will set up an ensign for the nations, and will assemble the dispersed of Israel, and gather the scattered of Judah from the four corners of the earth. (Isa. 11:11–12; trans. from Scherman 2000)

Furthermore, the prophet Zephaniah implies that Jews were living in Ethiopia through his statement: "From the other side of the rivers of Cush, My supplicants, groups of My scattered ones, will bring my tribute" (Zeph. 3:10; trans. from Scherman 2000). Unfortunately, the biblical record of the prophets is difficult to take at face value for studying history. Isaiah is believed to have lived during the eighth century BCE and Zephaniah during the seventh century BCE, but there is disagreement among scholars of biblical history as to the authentic authorship of these books in the Bible on whether or not these respective prophets did write the first editions of these books or if someone else did subsequently (particularly concerning Isaiah and a *Deutero-Isaiah*). Regardless of the

truth on the authorship of these prophetic books of the Bible, what is significant in relation to the argument for an ancient origin for the Beta Israel in Ethiopia is that Jews are mentioned as being there during this period in history on more than one occasion. It is thus not unreasonable to assume that a Jewish presence in Ethiopia was known in Israel at this time. Therefore, it is possible that this Jewish presence could have been the initial colonizing Jewish population in Ethiopia, who were subsequently joined by others over the centuries and, together, became the ancestors of the Beta Israel that we know today. During this period, the Jews, more commonly known as Israelites, comprised 12 different tribes. This Israelite group could have been one of the tribes (or a significant part of one), such as Dan, that was being dispersed across the ancient world due to the conflicts the kingdoms of Judah and Israel were having with the Assyrian and Babylonian empires.

Other Legends and Oral Traditions

Besides these early accounts from the biblical period there also exist ancient legends that report on other instances of Jewish settlement in Ethiopia. The oldest, in regard to chorological order, pertains to Moses prior to his leadership of the Jewish people, when he led them on the exodus out of slavery in Egypt to freedom in Israel. According to the legend, Moses resided in Ethiopia for 40 years as the leader of the local inhabitants during his self-imposed exile prior to returning to Egypt to begin the exodus. A group of Israelites from Egypt accompanied him on his sojourn in Ethiopia and chose to remain (Zegeye N/A: 10).

A second legend involves King Solomon and the queen of Sheba. According to the legend, part of the queen of Sheba's dominion included what is now Ethiopia. The queen of Sheba, captivated by the wisdom and reputation of King Solomon, visited him at his palace in Jerusalem. Their relationship involved more than platonic friendship (different variations on specific details of this legend exist), and they produced a son, Menelik I. The queen of Sheba and Menelik I eventually returned to northeast Africa, where Menelik I became the founder of the Solomonic Dynasty—a dynasty that ruled Ethiopia for centuries. However, the religion this dynasty practiced during the historical period was Christianity, not Judaism. A variant of this legend also claims that when Menelik I left Israel for Ethiopia, he smuggled out with him the Ark of the Covenant from King Solomon's temple as well as an entourage of temple priests (Isaac 2005: 125; Zegeye N/A: 6, 9–10).

King Solomon was of the tribe of Judah, and Moses and the temple priests were of the tribe of Levi, implying a migration of members of these tribes to Ethiopia in addition to the others discussed earlier. However, the tribes of Judah and Levi are two tribes that were not part of the 10 Lost Tribes. While provocative and interesting as legend and oral history among the Beta Israel and other peoples of Ethiopia, these stories are difficult to use as evidence to substantiate

the argument that the Beta Israel are descended from a Lost Tribe. Furthermore, these stories beg the question on whether descent from a single individual from the Jewish people is sufficient to claim descent from a Lost Tribe, particularly considering that the tribes of Levi and Judah were never among the groups that were "lost."

Conclusion

It is certain that the origins of the Beta Israel in Ethiopia are ancient and shrouded in mystery. The oldest information on a Jewish presence in Ethiopia is from biblical accounts and legends, which is considered less reliable as an objective historical document. Most of the stories that originated from ancient Israel and Ethiopia are not in agreement with one another; however, they do not discredit one another either. What they do have in common is that the earliest Jewish presence in Ethiopia, the ancestors of the Beta Israel, could have begun during the biblical period. The Judaism of the biblical period was far from the form that the modern world identifies as the Jewish faith, and in Ethiopia it developed almost independently from the rest of world's Jewry following the destruction of the temple(s) in Jerusalem. This raises the question of who the first Jews that came to Ethiopia during this ancient time in history were, as well as to the fates of the 10 Lost Tribes of Israel. While there does not appear to be sufficient evidence to unequivocally identify one tribe over another (lost or not) as the ancestors of the Beta Israel, there does exist a plausible theory that the Beta Israel are descended from the Lost Tribe of Dan based on the oral tradition and legends of the Beta Israel and the account of Eldad. Although the ninth-century account of Eldad is wrought with embellishment and can't entirely be taken at face value, there is some viable historical information that can be extracted, in a similar fashion as has been done with Marco Polo's *Travels* as well as with the application of oral history and traditions.

An example of an oral tradition with some historic merit found across many parts of the world concerns the Jewish people and the priestly clan of the Kohanim within the tribe of Levi (one of the non-lost tribes). For millennia, the Kohanim have continued to identify their descent and place among the Jewish people through an oral tradition that dates back to the biblical period. According to Jewish tradition, the founding male ancestor was the prophet Aaron, the first high priest. Recently, genetic studies of Y-chromosomal DNA have been conducted that corroborate the oral tradition of the Kohanim as coming from a single founding male ancestor (Boster, Hudson, and Gaulin 1998: 967–71). Similar genetic studies have been conducted on the Beta Israel, but no positive correlations for the Y-chromosomal DNA of the Kohanim have been found (Kleiman 2004: 83–84). This conclusion only excludes the possibility of descent from the Kohanim founding male ancestor. This

finding is significant because an oral tradition claiming descent from the Kohanim is not prevalent among the Beta Israel, whereas individuals who claim descent from this clan can be found throughout the Jewish diaspora. Due to our still limited knowledge of the human genome, scientists and historians still have yet to define the limits of when genetic testing is an appropriate method for confirming or disconfirming questions pertaining to migration patterns and genealogical studies in ancient human history. However, the oral traditions of the Kohanim among the Jewish people, which was later scientifically reaffirmed in the late 20th century to have some historical credit (with a single founding male ancestor), is significant as supporting the validity of some elements of oral tradition, even when they are ancient.

Therefore, due to the combined information provided by the historical record in Ethiopia and elsewhere, the biblical record, and the oral traditions and legends of the Beta Israel, it is not unreasonable at this time to conclude that the founding ancestors could have been from one or more of the 10 Lost Tribes of Israel and that this population could have been added to as described in the other historical theories proposed by this theory's cynics.

References and Further Reading

Adler, Elkan Nathan, ed. "Eldad the Danite (ca. 880)." In *Jewish Travelers in the Middle Ages: 19 Firsthand Accounts* (pp. 4–21). New York: Dover, 1987.

Barnavi, Elie, and Miriam Eliav-Feldon. *A Historical Atlas of the Jewish People: From the Time of the Patriarchs to the Present*. New York: Knopf, 1992.

Boster, James S., Richard R. Hudson, and Steven J. C. Gaulin. "High Paternity Certainties of Jewish Priests." *American Anthropologist*, New Series, 100, 4 (December 1998): 967–71.

Corinaldi, Michael. *Jewish Identity: The Case of Ethiopian Jewry*. Jerusalem: Magnes Press, Hebrew University, 1998.

Isaac, Ephraim. "The Question of Jewish Identity and Ethiopian Jewish Origins." *Midstream Magazine: A Monthly Jewish Review* (October 2005).

Jackson, Peter. "Marco Polo and His 'Travels.'" *Bulletin of the School of Oriental and African Studies, University of London* 61, 1 (1998): 82–101.

Kaplan, Steven. "Indigenous Categories and the Study of World Religions in Ethiopia: The Case of the Beta Israel (Falasha)." *Journal of Religion in Africa* 22, 3 (August 1992): 208–21.

Kessler, D. *The Falasha*. New York: Africana Publishing, 1982.

Kleiman, Yaakov. *DNA and Tradition: The Genetic Link to the Ancient Hebrews*. New York: Devora, 2004.

Lecker, Michael. "Judaism among Kinda and the Ridda of Kinda." *Journal of the American Oriental Society* 115, 4 (October–December 1995): 635–50.

Pankhurst, Richard. "The Falashas, or Judaic Ethiopians, in Their Christian Ethiopian Setting." *African Affairs* 91 (October 1992): 567–82.

Quirin, James. "The Process of Caste Formation in Ethiopia: A Study of the Beta Israel (Felasha), 1270–1868." *International Journal of African Historical Studies* 12, 2 (1979): 235–58.

Quirin, James. "Oral Traditions as Historical Sources in Ethiopia: The Case of the Beta Israel (Falasha)." *History in Africa* 20 (1993): 297–312.

Rosenberg, Stephen G. "The Jewish Temple at Elephantine." *Near Eastern Archaeology* 67, 1 (March 2004): 4–13.

Scherman, Rabbi Nosson. *Tanach, the Torah/Prophets/Writings: The Twenty-Four Books of the Bible Newly Translated and Annotated*. New York: Mesorah Publications, 2000.

Strack, H. L., and Günter Stemberger. Introduction. In *Talmud and Midrash*. Trans. and ed. Markus Bockmuehl. Minneapolis: Fortress Press, 1996.

Weil, Shalva. *The Religious Beliefs and Practices of Ethiopian Jews in Israel*. Hebrew. Jerusalem: NCJW Research Institute for Innovation in Education, Hebrew University of Jerusalem, 1991.

Zegeye, Abebe. *Narrative Discourse, Social Crisis and the Construction of the Beta Israel Identity*. University of South Africa, N/A.

CON

The debate regarding the authentic Jewishness of the Beta Israel (Falasha) people of Ethiopia, since the 1984 airlift of 10,000 members of the tribe from outside the Gondar province of Ethiopia to Israel, has persisted unabated to our present day. Perceived as *genuine* members of a lost tribe of Israel, Beta Israel (House of Israel) tribe members, as the Lemba people in South Africa, were later proven through genetic tests to be direct descendants of the Cohanim (direct bloodline to Old Israel's priest class), after a sizable portion of its membership had broken away from Israel, around 931 BCE, shortly after King Solomon's death. As the Lemba tribe members, Beta Israel tribe members have also always been convinced, even *prior* to their airlift to Israel in 1984, that they were one of the 10 Lost Tribes that had broken away. They have been supported in their claims by numerous African American Jewish groups, and gradually, by fellow Israelis and global members of the Jewish community. Other members of Israeli society and the global Jewish community, many of whom are of Ashkenazi (European) parentage possessing the racist preconceptions acquired from those former lands their ancestors had lived in, have been discriminated against and annihilated in

and during the alienating diaspora period. They had resettled in Israel and ironically used these same stereotypical preconceptions that had been earlier burdened within their history and applied them against fellow brethren members of the Jewish faith, who had just rejoined Israeli society after their own brutal and languishing diaspora period. This hostility toward foreigners, or other nonsimilar human groupings, has been historically experienced in the political, social, cultural, and moral settings of most societies in varied multihued layers. Except in this case, the only difference, or so it seems, is that many members of the Ashkenazi group within Israel—who have been considered the more dominant sect in society since independence in the late 1940s, to which they would never admit as the real motives behind their individual or collective perceptions—do not "seem" to be "real" Jews. That, to many Jews and non-Jews, would be perceived as tantamount to some Jews not living up to the ideals to which modern Israeli society had initially been created. For most, if not all members of Israeli society, such references would be regarded as code words for the fact that Beta Israel members were singled out as blacks "claiming" to be Jews; instead of accepting the fact that they were *real* Jews. This leads to a much more profound question as to what, or even more important *who*, within the Jewish faith defines or represents to *whom* within the faith what constitutes a *real* Jew? The important *identity* issue has continued to challenge Israeli Jews, as their history unfolds, in regard to their innate self and collective perceptions and assumptions about their society. The historical facts and sources, as will be shown in this section, do not seem to unequivocally support those proponents who believe they're not a Lost Tribe.

In accordance with early biblical texts, with the death of King Solomon in 931 BCE and the ascension to the throne of his son Rehoboam, 10 of Israel's northern tribes broke away from the greater Israel, which they had earlier identified with and given allegiance to. They followed their new king, Jeroboam. Two new territorial entities were created within Israel. One would be located in the north, the other in the south. The only two Jewish tribes remaining that followed the house of David were the tribes of Judah and Benjamin. The other 10 tribes faithfully followed Jeroboam. The tribes of Judah and Benjamin were located in the southern area and became known as the kingdom of Judah (obviously from the name, the tribe of Judah predominated the tribe of Benjamin). The other 10 tribes controlled the northern area under King Jeroboam's rule, despite his lack of legitimacy as a ruler of any greater Israel entity, since he was not in the accepted royal line of succession in the house of David. The dominant tribe in the north was Ephraim, like Judah in the south, which had held its sway in the naming of the new political entity (the 10 tribes to the north named their political entity Ephraim). After two centuries of separation, some members and tribes from the north, as the Levites, returned to the southern portion. According to those early biblical texts, this return to Judah was pursued because King Jeroboam and his successors had turned their nation into that of idol worshipers.

Besides Ephraim, the tribes of Simeon and Manasseh would also later return to Judah. By 722 BCE, the Assyrian Empire to the north of the Israeli kingdom invaded and conquered it. The Assyrian king Shalmaneser V ordered that Israel's inhabitants be exiled, after a few revolts by its Jewish inhabitants. They were forced out of their land into what is present-day Syria, Iraq, and Iran. While this chaos ensued, by 716, other northern Jewish tribes returned to the Judean kingdom, which was ruled by King Hezekiah. In the book of II Chronicles of Bible, it is stipulated that King Hezekiah declared to the Jewish inhabitants of his Judean kingdom that as children of Israel, after this protracted intratribal infighting period, all dispersed, returning, and Judean Jews alike, in a new and reinvigorated unity, needed to "turn again unto the Lord G-d of Abraham, Isaac, and Israel," and only if they did so, then G-d "will return to the *remnant* of *you* that are escaped out of the hand of the Kings of Assyria" (II Chron. 30:6, emphasis added). Those were very powerful words that would return to haunt and test the new Jewish national entity two millennia later. In accordance with that text, despite the people's dispersal, it did not in any way imply that they were lost or forgotten by God. So, under that pretext, if the Lord God doesn't forget his children, as affirmed since the beginning of creation through the *eternal essence* of their individual and shared humanity, then what man or woman, despite whatever misguided notions of "superiority" or "exclusiveness" they might strongly wish to believe in and impose on others, can cast God's words asunder?

Living in complete ignorance of Africa and its interior until the mid-19th century, European nations and their subjects relegated Africa's inhabitants to the same predicament that they had earlier done with North America's Native Americans. In each of the European capitals' foreign offices, Africans were simply not regularly factored into the "grander" imperial designs and schemes drawn up by them, their civil servants, and proxy "adventurers" in the never-ending quest to penetrate and control the treasures, overland trails, and waterways within the African continent. An insidious sense of "whiteness" during that period, despite the abolition of slavery in England earlier in the 19th century, and Europe, and later in the United States, the foundation of racism held its economic and psychological sway on many white citizens, who felt they could benefit from the privileges often bestowed upon them in their subtle and often not-so-subtle continued identification with their race. The archaeological, social, and cultural interpretations of the period, which usually were dominated by a racist European scholarly preorientation, especially in relation to recently conquered societies and tribes, invasively persisted to regularly imbue all whites with a sense of "superiority" in racial discourses that would unfold with the recently subjugated local inhabitants. The insidious aspect to this preconstructed discourse introduced by the European powers and their representatives would infuse this illusory sense of superiority to working-class whites, who often were and felt marginalized by unfolding economic, social, cultural, and class differences

and realities they suffered under within the larger settings of their *own* societies. For being "white" in relation to the newly conquered and acquired territories and societies, along with their diverse local inhabitants within the global context, provided and secured for them, along with their European, more educated, elite countrymen, a social deference and rights that most, if not all, the conquered local inhabitants were not permitted to enjoy. In such global ventures, the working-class white populations of Europe and the United States solidified their bonds at any opportunity with their fellow white countrymen, who identified with a more elite class than they had been able to. They put aside any earlier labor, social, and class grievances they might have held toward members of their own country's elite class. According to them, it was all for the public and global racial good. It morphed and became synonymous with the old and often well-referenced metaphor of having two (very comfortable) peas in a pod. Of course, that played out at the expense of the recently conquered colonies and their diverse local inhabitants. It's within that specific context that the existence of the Beta Israel tribe would be discovered in the late 19th century.

The Beta Israel tribe had existed in the Gondar province of the northwestern Ethiopian highlands, centuries *prior* to them playing any public role in Ethiopian society at the beginning of the 17th century. The term used for them by Ethiopians prior to the 14th century and into the early 17th century was *ayhud* (Arabic), which was later adapted into the Amarinia (Amharic) language for Jews. In an attempt to later marginalize their presence and influence in society, they were called Falasha (which meant *strangers* or *wanderers* in Amarinia) by the surrounding non-Jewish population, which members of the Beta Israel tribe considered derogatory and recoiled at its use to describe them. Then, shortly thereafter, during the Gondar dynasty, which ran from 1632 to 1755, Beta Israel tribe members acquired work as artisans and tenant farmers. From 1632 until the mid-18th century, Beta Israel settled in the urban areas and evolved into critical and vibrant members of the larger Ethiopian Amarinia and Tigre-speaking community. Elites within the Beta Israel tribe gained enough favor that they were awarded significant land grants. It helped that they were perceived by fellow Ethiopians, who were all primarily of the orthodox Ethiopian Coptic religious persuasion, as being very skilled carpenters and masons who played a central role in the construction of local and regional palaces and churches.

But, from 1755 to the late 19th century, Beta Israel, Muslim, and other non-Christian minorities were singled out by the larger Ethiopian Christian community as contributing to the economic and social decline and rampant state of famine that they individually and collectively experienced. General hearsay and Ethiopian folklore, with no documented proof available to scholars and historians, asserted that as a renegade Christian tribe, Beta Israel took up arms against the majority of decent orthodox Christian Coptic Ethiopians and lost. Thereby, to supposedly spite the majority, they embraced Judaism. Once they did that,

according to local lore, the majority of Ethiopian Christians took their lands and left them landless, forcing them to enter the lowest of the artisan trades in order to survive economically. This had the result of radicalizing the Beta Israel people, alienating them from Ethiopian society, and causing them to seek a new affiliation. It was, of course, a very convenient scapegoating technique used by a dominant majority against its vulnerable minorities during times of hardship, as one often sees used in most, if not, all societies and human gatherings throughout world history. Besides the brown shirt youth and gestapo tactics of Hitler's early years in power as Germany's chancellor in the early 1930s, with its confiscation of Jewish and fellow minorities land and extensive property, a similar incident that infamously stands out in U.S. history during World War II was not only the detention of Japanese Americans in labor camps, but also the excuses creatively and despicably conjured up by former neighbors of the interned Japanese American citizens and town officials that their individual and collective legal ownership of land was to be confiscated and sold to more "patriotic" Americans. The dominant majority in most of these historical contexts was not motivated by any sort of *genuine* patriotism, racism, or even religious intolerance; though sometimes, as history has regularly shown, it was convenient for the instigators or culprits of such deeds, to refer back to them by exhibiting no regrets or serious qualms of guilt as an excuse. It was cold-calculated minds and motives that reigned supreme. They were simply motivated by raw and jaded greed, and of course, the ever important bottom line. That has often been the sad standard during such wrenching historical moments.

The Gondar province was also often invaded by the Mahdist Muslim Sudanese forces, which of course, didn't help much in the alleviation of the harassment and discrimination against marginalized minorities as Beta Israel, the Muslims, and other non-Christian minorities. With the later arrival of Protestant European missionaries, matters didn't improve much for Beta Israel. The resentment and persecution they endured during that period were only further aggravated with the unique home-bred brand of tunnel-vision intolerance that most European missionaries brought to the African landscape in general, and within Ethiopia's northwest highlands in particular. The swill of that new imported brand of persecution was especially directed at the Beta Israel tribe, since most members of the tribe refused to succumb to European Christian conversion efforts. It was only later, following the reestablishment of contact with members of Western Jewry, that some of the harassment and discriminatory practices imposed on the Beta Israel tribe, with the increasing widespread momentum of religious conversion campaigns at the hands of European missionaries in conjunction with the official support of political representatives of European foreign powers, began to slowly recede.

The Lost Tribe perspective at that juncture began to reassert itself and grew stronger within the ranks of the tribe members. They strongly believed that they

were direct religious and ethnic descendants of the earliest Jewish migration into Ethiopia. It was the *authentic* Judaism, according to them, that their ancestors brought as they traveled from Jerusalem to Syria, through the Arabian Peninsula, into Yemen, and finally settling in Ethiopia. They were not merely the direct descendants of King Menelik I, the son of King Solomon and Makida, the queen of Sheba, who most, if not all, Ethiopian Christians attribute their royal families' origin to over the past two millennia; but, instead, they were authentic members of a lost Jewish tribe that migrated through the Arabian Peninsula and Yemen and finally settled in Ethiopia.

Menelik, son of the Queen of Sheba, standing before his mother who is holding a mirror. Illuminated page from *The History of the Queen of Sheba* **(Ge'ez and Amharic). Ethiopia, produced in the late 19th century. (Bildarchiv Preussischer Kulturbesitz/Art Resource, NY)**

The Celebrated Search for the Lost Ark of the Covenant

One of the regular stories concerning biblical-times-Ethiopia surrounded the legend of the Ark of the Covenant, and that it might have survived in Ethiopia. The legend was that King Menelik I, believed by tradition to be the son of the queen of Sheba, brought it back with him to Ethiopia after visiting his father, King Solomon. This legend has led to many books, magazine articles, and television documentaries. The most famous of these was the entirely fictional film *Indian Jones and the Raiders of the Lost Ark* (1981).

Certainly members of the Ethiopian orthodox church claim to have the Ark of the Covenant in the Church of Our Lady Mary of Zion, in Axum, where it has been used for a range of ritual purposes. The original church on the site of the present one was probably built in the fourth century, with Emperor Haile Selassie and the Ethiopian imperial family paying for the construction of a new chapel there.

On the other hand, a counterargument with a "convert" interpretation makes the claim that members of the Beta Israel tribe were originally Christians who had not experienced a mass migration as the Lost Tribe idea entails. Instead, the "convert" argument supporters believed that this local Ethiopian Christian sect might have been influenced by a few Jewish migrants into

Ethiopia over the past millennium or so; but within their hearts they *knew* they had adopted some Jewish practices, so that they would stand out in relation to other Christian sects within Ethiopia. By doing so, members of the Beta Israel tribe had hoped to seize the higher moral ground or religious standard that they knew their fellow Christian believers would be unable to claim, despite the fact that they collectively knew that the earliest Christians owed their existence and teachings to the influence of the Old Testament prophets and their people. Therefore, Ethiopian Christians strongly believed that instead of being direct descendants to one of the 10 lost tribes of Israel, the earliest members of the Beta Israel tribe were direct descendants of these earliest Ethiopian converts from Christianity to Judaism. The local Christian inhabitants were definitely not going to be out-notched in their most basic beliefs and values by a group of fellow "Christians" just because they took it upon themselves to reembrace Judaism. This would feed into a more renegade or rebel interpretation as to the constructs of the Beta Israel tribe. That goes much further into the former Christian argument and unabashedly claims that not only were the ancestors of the Beta Israel tribe former Christians, but Christian influence saturated the *core essence* of the Beta Israelis faith. According to this argument, all beliefs and practices of members of the Beta Israel tribe can be adduced to a common Christian origin.

That, therefore, should finally bring one to question the motives and *intrinsic* hatred of the dominant majority group toward Beta Israel during most, if not all, of its earlier historical periods of brutal discrimination of the Beta Israel tribe. Was it simply a matter of sour grapes *within* the Ethiopian national *psyche* ("How could members of the Beta Israel tribe be calling themselves the *real* Jews when in reality everyone should know we're the "real" Jews?"). Ethiopian Christians strongly felt that Beta Israel had stolen their thunder. This might explain why in the past two millennia most of Ethiopia's Christians felt they needed to subvert the *real* set of facts and history of Beta Israel, regularly infusing it with their own unique brand of disinformation, throwing up regular smoke screens to distract and deflect interested scholars and outsiders from gaining clear insight into what the historical evidence and facts actually were. Like the Ashkenazi Israelis and their identity crisis as to what defined a Jew, the dominant Ethiopian Christian Coptic group had a very difficult time of coming to terms with their own *identity* as early Christians, let alone as Ethiopians. In both cases, it appears that scape-goating the Beta Israel tribe to somehow compensate for their individual and collective shortcomings was, and continues to be, a very shallow way of dealing with such profound issues. When the objective reader follows those threads, specific patterns, beliefs, value systems, and practices stand out. First, there's no archaeological or historical documented evidence that in less than 300 years, from the early 15th to the late 16th centuries, any *ayhud* group, or mix of united groups, emerged to battle their Christian Ethiopian brethren. Second, there's no evidence, written or otherwise, that

stipulates that the *ayhud* were a heretical Christian sect. The only evidences that have survived are the oral traditions and Jewish daily religious practices performed by all members of Beta Israel—like that of the Lemba Jewish tribe in South Africa. They only worshiped Adonai, the God of Abraham, Moses, and Isaac. They performed circumcisions, held to Shabbat, and ate kosher food. They withstood horrific suffering at the hands of their Ethiopian Christian neighbors. Their tenacity was fierce in their pitched battles with Ethiopian Christians. Their individual and collective readiness to die for their faith grew legendary in local Ethiopian folklore. In one such battle, a Beta Israel woman jumped off a high cliff, holding on fiercely to one of her Christian tormenters and pulling him down with her, beseeching Adonai to save her. With that kind of readiness to perish and martyr oneself for one's beliefs, which according to their earliest traditions was permissible, so that God's name might be sanctified, *al Kiddush ha-Shem*; and what fierce fighting and sacrifices had ensued from their ancestors in their battles against Greece's forces, while the latter were attempting to reach out for a larger swathe of the Mediterranean World; it's, therefore, very hard to believe that that woman or members of her tribe were recent converts from Christianity as Ethiopian Christians would have others believe. It would have been much easier for the woman or members of Beta Israel to reembrace Christianity, instead of so obstinately holding on to what they very strongly believed was the faith of their ancestors, who made that long trek, centuries earlier, from Jerusalem. It was hardwired into the deepest recesses of their souls and spirits. That is, they were Jews.

In the status of continued limbo during that unique period of diaspora that Beta Israel experienced, readers should also consider that during much of the period which had spanned almost two millennia, they didn't have the comforts of present-day technology or access to 24-hour nonstop cable news. They couldn't have easily been influenced by the latest fads in intellectual, religious, or resistance to power logic (i.e., talking truth to power); instead they held very tightly to traditions and practices handed down to them by their ancestors over those millennia. To Beta Israel, those traditions and practices were as (and in most cases more) important, for they strongly believed it was given to them by God, as if one had received a family heirloom, and they were told by their elders to hold on to it, treasure it, and then pass it on to the next generation(s). Beta Israel had grown very mistrustful of their neighbors (can't say we can really blame them, with the kind of reception they received) and withdrew into a more monastic lifestyle, which was very much in touch with the practices and traditions of their earliest ancestors. Many detractors of Beta Israel have attributed their monastic lifestyle to practices they took with them, or became influenced bythem, from their Ethiopian Christian neighbors. That argument appears to fly against traditions in all old Jewish communities, as in the old Jewish sects of the Essenes and Therapeute. Historical records show that many of their

inhabitants lived in celibacy, while retaining their independence from surrounding Jewish communities, often living monastic lives. These were very old, established Jewish practices from years *prior* to the arrival of Jesus of Nazareth and later evolving small Christian communities in Jerusalem and beyond. Christianity didn't introduce monasticism to the Jews, it was the other way around.

Over 20,000 Beta Israel tribe members experienced their own unique exodus from Ethiopia in 1977 to 1985, which could be equated in the scope of its trauma to the biblical exodus from Egypt. They set out on foot to the southern border of Sudan, following another wave of persecution instigated this time by a whole different set of circumstances. Ethiopia's new political and economic system emerged into an Ethiopianized secular Marxist version of the ailing Soviet model. Mengistu Haile Miriam became the country's new leader, after years of bloody succession intrigues, which followed Emperor Haile Selassie's murder in his bed in 1974. Now it wasn't only the religious Ethiopian Christian locals that had Beta Israel in their sights, but also the new Marxist junta that was out to eradicate all traces of religious practices throughout the country—including the Beta Israel variety. Beta Israel picked up their belongings and never looked back. The trip from Ethiopia to Sudan was fraught with all kinds of dangers. Many Ethiopian locals victimized them and attempted, during consistent intervals, to intimidate and regularly rob them of their possessions during the long trek. They suffered from seriously dwindling food, water, and medical supplies. But a lack of shade or cover from the merciless heat of the day seemed to be one of their major obstacles. It is believed that Beta Israel lost around 4,000 members of the original 20,000 who had set out on the trip back to their perceived homeland, Israel. When they finally reached Sudan, they were placed in hastily constructed camps, where they were isolated and in some cases mistreated by the locals. They remained in worsening wallowed conditions for two years. The fortunate ones were airlifted a year after their arrival. Their mettle had been tested. The only thought that seemed to sustain them during this very difficult time was that they were returning to the Promised Land, land of their forefathers and -mothers.

But once they had arrived, they experienced the genetic and pseudo-scientific racism that thrived in the late 19th and much of the 20th centuries in Europe and the United States. It was and continues to be so invasive, within certain Jewish groups within Israeli society, as many of the Ashkenazim, that it poisons the discourse between members of Beta Israel and some of the larger Jewish groups within and outside of Israel. Instead of showing empathy to Beta Israel for the horrific experiences they endured during their diaspora over the past two millennia, that most, if not all European Jews, had experienced in Western Europe with the Spanish Inquisition, the Holocaust, and the daily slights with the usual painful slings and arrows that would grow into blatant and violent anti-Semitism, they ironically singled and continue to single out Beta Israel, fellow Jewish brethren *within* the faith, as they had been regularly

singled out during their shared experiences during their European and American diaspora by non-Jews. Quaint dismissive comments of some of the Ashkenazi variety have questioned that if they were *real* Jews, how come they're unfamiliar with the Talmud and do not practice regular Jewish dietary laws. Innocuous comments persisted about Beta Israel: "They don't understand or speak Hebrew. They didn't know regular Jewish practices and festivals, as *Shavuot* (the Pentecost), *Purim*, and the New Year for Trees, *Tu Bishvat*." What many of those subtle and not-so-subtle dyed-in-the-wool racists consciously ignore are the historical facts, that the Talmud and all the shared dietary laws, daily practices, and festival recognitions were consensually agreed to during varying reform movements that arose during the Jewish diaspora period in Eastern and Western Europe. At that time, Beta Israel was in total isolation from much of the known world. That's why they're referred to as a "lost tribe." They "lost" and had become completely "disconnected" from the unfolding Jewish history, outside in the world. Their ties inadvertently had become severed. The most important reality they tried as hard as they could, even under repressive circumstances, to hold on to for over two millennia was their bond with their God. The fact that the Judaism they brought to Israel did not *mesh* exactly with the Judaism being practiced by the majority did not make them any less Jewish. It is similar to cases one reads about in the current news where a child was lost to his parents and was reunited decades later with his family. The child's memories will not be similar to the shared memory of his or her biological and cultural family. Does that make that child in any way less their child then when he or she lost touch with them during childhood? In such matters, special circumstances and details need to be taken into consideration so that the child, or adult for that matter, is happily reunited with his or her parents and siblings. Each Jewish individual and group has had their own *unique* experiences to share as Jews with the larger extremely diverse Jewish whole within Israel. Polish Jews brought parts of Poland and their experiences, negative and positive, with them. The same can be said for Russian Jews and all other Jews who had settled, and continue to settle, Israel from the most far regions of the world. No one ever questioned their *Jewishness*.

Why then should they start with Beta Israel? Judaism, as most dedicated and conscientious Jews have always recognized, was and is a religion and a set of moral precepts to be daily practiced and reformed, if need be. It was never and will never be a race. So, the important question that needs to be asked is why, at this time in a more enlightened and sophisticated 21st century, are there still intelligent Jews in Israel questioning the initial *Jewishness* of a lost tribe within their midst, as many other members of other lost tribes from the original 10 tribes that left and then returned to Israel, never had their *Jewishness* ever questioned? They all had to learn the Hebrew alphabet from scratch, as Beta Israel has had to do. All resettled Israelis, by legal and common human

decency, would have to emphatically include Beta Israel in the shared birthright to a promised land within the larger Israeli entity. For if it is just blatant racism, in its ugliest and most corrosive form, as it clearly seems to have become for a few sad sorts within Israel, it is betraying within their individual *psyche* the historical bondage they and their families suffered while living in the diaspora under what they justifiably attributed to be the perverse nature of many non-Jewish pogroms and oppression; but, they later found themselves in their resettled lives in Israel living under the rubric of Greece's Homer, who was quoted by Plato as saying, "If you make a man a slave, that very day / Far-sounding Zeus takes half his wits away." With half their wits gone, then one might be able to understand why most of these troubled sorts betray a lack of general historical knowledge: specifically, their own rich Jewish history as to what, if anything, defines "whiteness" (if there is ever an intelligent definition for such a term, as hard as many specialists and scholars have unsuccessfully sought to do over the centuries, or for that matter for any other race) and "Jewishness." Many of those who would be willing to marginalize the Beta Israel from the mainstream of Jewish life, as practiced in Israel and throughout the global Jewish community by mixing *Jewishness* with *whiteness*, needs to remember those historical lessons. There's a better than odd chance, being the robust and strong-willed Semites that they were in the part of the Mediterranean world within which they resided, that the prophets Noah, Abraham, Isaac, Jacob, Joseph, Moses, David, and Solomon and their people were *not* white, biologically, culturally, or otherwise. Besides most likely possessing dark features and complexions, they *all* saw themselves as Jews. What they shared with Beta Israel is that they were all Jews. They identified with their Jewish faith and its strong bond with God and sought to meticulously cultivate their *essential* shared *Jewishness* as a chosen people, instead of only perceiving themselves as a chosen people. Race never was a factor. That, according to them, was not the most humane decent Jewish prism through which they needed to view such matters. They and many of their followers were busy trying to be decent, caring, loving, humane Jews. Therefore, in regard to Beta Israel, that's what some of these few discriminating Jews within Israel and beyond need to seriously consider—not only that they need to be decent humans in the eyes of God, his prophets, their communities, and families, but, most importantly, *within* their individual selves. For, by consciously and conscientiously doing so, they will be able to live up to the highest ideals of their ancient people, and that is to be *good* for their individual selves, toward one another, and ultimately toward the Lord.

References and Further Reading

Bard, Mitchell G. *From Tragedy to Triumph: The Politics Behind the Rescue of Ethiopian Jewry.* Westport, CT: Praeger, 2002.

Chidester, D., and Robert P. Scharlemann. *Savage Systems: Colonialism and Comparative Religion in Southern Africa*. Charlottesville: University of Virginia Press, 1996.

Elazar, Daniel. *The Other Jews*. New York: Basic Books, 1989.

Kaplan, Steven. *Beta Israel (Falasha) in Ethiopia: From Earliest Times to the Twentieth Century*. New York: New York University Press, 1995.

Quirin, J. *The Evolution of the Ethiopian Jews: A History of the Beta Israel (Falasha) to 1920*. Philadelphia: University of Pennsylvania Press, 1992.

Renard, Raymond. *Sepharad*. Mons, Belgium: Annales Universitaires de Mons, 1966.

Robert, Dana L. *Brief History of Missions*. Oxford: Wiley, 2009.

Rodrigue, A. *Images of Sephardi and Eastern Jewries: Transition: The Teachers of the Alliance Israelite Universelle, 1860–1939*. Seattle: University of Washington Press, 1993.

Sacerdoti, Annie, and Luca Fiorentino. *Guido all'Italia ebraica*. Casale Monferrato: Marietti, 1986.

Shelemay, Kay Kaufman. "Continuity and Change in the Liturgy of the Falasha." In *Modern Ethiopia: Proceedings of the Fifth International Conference of Ethiopian Studies* (pp. 479–87). Nice: 1977.

Shillington, Kevin. *History of Africa*. New York: Palgrave Macmillan, 2005.

Stern, H. A. *Wanderings among the Falashas in Abyssinia*. London: Taylor and Francis, 1968.

Tibebu, Teshale. *The Making of Modern Ethiopia, 1896–1974*. Lawrenceville, NJ: Red Sea Press, 1995.

Wilson, Marvin R. *Our Father Abraham: Jewish Roots of the Christian Faith*. Grand Rapids, MI: Eerdman, 1990.

Zafrani, Haim. *Mills ans de vie juive au Maroc*. Paris: Maisonneuve et Larose, 1983.

13

Ancient findings of Ancient Babylonian cities confirm the Old Testament.

PRO   Benjamin D. Thomas
CON   Thaddeus Nelson

PRO

Professor Friedrich Delitzsch, eminent German Assyriologist, was delivering a series of three lectures starting on January 13, 1902, on the recent findings of cuneiform research in connection with the Bible before the newly established German Oriental Society under the sponsorship of the German emperor Kaiser Wilhelm II. A great stir arose among the populace in anticipation of Delitzsch's address. He was the son of the distinguished and admired German Old Testament professor Franz Delitzsch, the latter being quite conservative in his views on the Bible, a devout Lutheran, and even a sympathizer of modern Judaism. As a result, the attending audience expected Friedrich to confirm their traditional, orthodox views regarding the Old Testament (or Hebrew Bible).

In his first lecture the German professor, though more reserved in his initial address, argued that Babylonian culture and religion ("Babel") were superior to the culture and religion characterized in the Old Testament ("Bible"). What's more, Babylonian culture directly governed and even permeated biblical culture. In Delitzsch's opinion, it was obvious that the connection between the Old Testament and Babylonian culture was not a matter of simple comparison, but that Babylon itself was the ultimate source of Old Testament thought and religious practice. He contended that the name of Israel's national deity, Yahweh, was already present in tablets from Babylon dating to the 2500 BCE (certainly too early!). Delitzsch regarded this as proof that monotheistic religion had been practiced by nomadic tribes in Babylon 1,000 years before the Israelites conceived of monotheism (according to the conventional dating of Delitzsch's time).

Delitzsch's deep antipathy to the Israelite religion, as a consequence of nurturing a growing infatuation with German fascism, was hinted at in the initial lecture when he suggested that the wife of the famous Assyrian king, Assurbanipal, was "a princess of Aryan blood." These and other claims were later discredited by notable scholars in the field of Old Testament studies.

In the second lecture, presented on January 12, 1903, Delitzsch went even further in his extreme views on the Old Testament. Three themes were

295

Akhenaten and Monotheism

A recurring theory in biblical scholarship is that the religion of the Pharaoh Akhenaten inspired the monotheism of the Israelites. There are two ways it could do so: either Akhenaten could have been the pharaoh during the time when the Hebrews were enslaved in Egypt, or his religion could have survived in some form so that they could encounter it later when they were so enslaved. Neither is well supported by what we know of Egyptian chronology, but Akhenaten is a mysterious figure, capable of surprise.

Akhenaten was the founder, and perhaps only true member, of the religion of Atenism: the worship of the Aten, the disc of the sun. Like Abraham, Akhenaten and his immediate family had a highly personal relationship with his deity; like Hebrew monotheism, Atenism eschewed the worship (and even the validity) of any other deity; and like Yahweh, the Aten was more vast and unknowable than the human-seeming petty gods of many polytheistic pantheons.

salient: (1) the denial of the "inspired" nature of the Old Testament; (2) a continuation of the ethical superiority of Babylon religion and culture to Hebrew religion and culture; and (3) a critical appraisal of Israelite monotheism.

Delitzsch assailed the Christian doctrine of "inspiration," which claims that God was the impetus behind the human authors when they wrote his message in the Bible. According to him, no divine revelation was contained in the Old Testament; rather, it was merely composed of the human will, whose ultimate derivation could be found in Babylon. Whichever Old Testament doctrines had been previously regarded as unique, Delitzsch attempted to demonstrate as common to the Babylonian worldview, which he always depicted to be the superior of the two. The second lecture, then, showed a marked contrast from the previous one. Shifting from comparative depictions between Babylon and the Bible, Delitzsch had begun an all-out assault on the importance of Old Testament religion.

It is helpful to step back momentarily and view these events from the perspective of their own historical background. Most of what Delitzsch was alleging was not new to the scholarly and public circles of his day. Indeed, a fraction of the German school of critical-historical theology had already shown signs of hostility to modern Judaism and the Old Testament. Moreover, in hindsight, Delitzsch acted as one might have expected, namely as a defender of Assyriology, which for many years had been relegated to an ancillary discipline of Old Testament studies.

The third and final lecture was delivered on October 27–28, 1904. Since his second lecture, Delitzsch had expressed orally his personal opinion to Emperor Wilhelm that he no longer believed the Old Testament was the inspired word of God. The emperor, in response, felt it necessary to issue a brief publication affirming his conviction in the divine inspiration of the Old Testament. Now, in

this final lecture, the emperor was not present nor was the audience of the German Oriental Society enthusiastic. (Nevertheless, Delitzsch did continue to have close relations with the emperor and the German Oriental Society for years thereafter.) Within the content of this final lecture, the focus of which was to make comparisons with the New Testament, the professor came dangerously close to saying that Jesus himself was not of Jewish, but of Aryan, stock and asserted, moreover, that many of the concepts held by Jesus found their origin in Babylon.

In subsequent publications by Delitzsch, what had been revealed only in part was now disclosed in full. He discussed in *Die grosse Taüschung* (The Great Deception) the opinion that the Old Testament essentially was the "Great Deception" and that much of it could be falsified or demonstrated to be primitive. In its place, the professor recommended that Wilhelm Schwaner's *Germanen-Bibel* should be esteemed. Now, furthermore, he openly attacked Judaism, declaring that it comprised a dangerous and impoverished people. He claimed that Jesus was only a newcomer to Judaism, not Jewish per se.

This section will outline some of the specific evidence that has been most persistent and influential in the Babel–Bible controversy and will defend how the findings from Babylon (or better, Mesopotamia) serve to confirm and authenticate the Old Testament.

The Evidence: The Creation Story

The Mesopotamian version of the creation story is told in the famous *Enuma Elish* (Akkadian meaning "when on high"). Even though the main deities differ from the Babylonian to biblical accounts—*Enuma Elish* with the god Marduk and Genesis 2, with the god Yahweh (not chapter 1)—there are several parallels between the Babylonian and the biblical creation stories. They are as follows: both refer to the watery beginning of Earth, which is then separated into the heavens and the earth; the terms for the water are etymologically related; the creation of light precedes the creation of the celestial bodies; the number seven figures prominently. The creation of man is the final act of creation and bears importance in both accounts. On the basis of these parallels, Delitzsch felt it necessary, in *Babel and Bible*, to state that "Indeed, the Old Testament poets and prophets even went so far as to transfer Marduk's heroic act directly to Yahwe" (1903: 49).

Several significant variances, nevertheless, may be observed between the two accounts. In the Babylonian creation story, the watery beginning of Earth is more specifically the goddess Tiamat ("ocean"); in Genesis, the water is inanimate, that is, not a living goddess. Moreover, whereas the gender of the Akkadian word *Tiamat* is feminine, the Hebrew word of Genesis (*tehom*) could be either masculine or feminine. Recently it has been contended that a direct philological borrowing of the Akkadian word for "ocean" into the Hebrew account is impossible. In the Babylonian account, Marduk uses the already existing body of Tiamat to form

the heavens and the earth. According to the traditional interpretation of Genesis, Yahweh created the heavens and the earth from nothing (*creatio ex nihilo*).

Under the influence of broader ancient Near Eastern culture, polytheism figures prominently in *Enuma Elish*. On the contrary, a singular deity, namely, Yahweh, plays the important role of creator in Genesis. The luminous bodies as enumerated in the Babylonian account were created in the order of stars, moon, and sun. Genesis, in contrast, follows the order sun, moon, and stars. Moreover, the concept of gates in the eastern and western horizon is lacking in Genesis. At present, no reference to the creation of animals or vegetation in any part of *Enuma Elish* has been found.

The creation of man, while finding significance in both stories, also has major differences. In the Babylonian account, man is made for the sake of relieving the gods from the necessity of providing their own food. Creation likewise is for the sake of the gods. In Genesis, while man is similarly made to serve God, creation is for the sake of man, who was to be its lord. The luminaries were, in the same way, made to benefit mankind in the biblical creation story. The Babylonian account is from the perspective of the gods; the biblical account from the perspective of man.

Conspicuously absent from the Babylonian version is the "Fall of Man" as it is told in Genesis 3. Some attempts have been made to find Babylonian archetypes of the fall of man in other texts, but all attempts have failed to convince. Again, the religious institution of the sabbath, as it was established on the seventh day of creation, finds no counterpart in *Enuma Elish* or in the larger religious institutions of Mesopotamia (as Delitzsch first proposed).

In the recent opinion of scholars, there is a definite reluctance to concede that any part of Genesis was directly derived from Babylonian parallels. Most say that there was no direct borrowing taking place from one account to the other. Some scholars even contend that even the most indirect borrowing is still beyond proof.

The Flood

The most remarkable parallels between Old Testament and Mesopotamian literature may be observed in the comparison of the flood narratives in the biblical book of Genesis with the famous *Gilgamesh Epic* and in the related story of *Atrahasis* (both titled after the names of the heroes, who are counterparts of the Noah of Genesis). Common to these stories is that the hero is instructed to build a boat with pitch. He loads his family and animals aboard. Both traditions agree that the boat landed on top of a mountain. In each, there is a bird scene in which one bird is sent out after another to search for dry ground. The raven and dove figure in both traditions. The hero offers a sacrifice to (the) God(s), who then blesses the hero. These and other parallels prompted Delitzsch to conclude that the story, as it was written down, was transmitted to Canaan. Many scholars

writing after Delitzsch believed this provided indisputable proof that the Hebrew authors of Genesis had borrowed *directly* from the Babylonian flood accounts.

Notwithstanding their basic commonality, there are considerable discrepancies between the Mesopotamian and biblical flood narratives at almost every level. The terms commonly translated as "boat/ark" are linguistically unrelated between Hebrew and Akkadian. The type of craft in *Gilgamesh* appears to be a cube, fashioned more after a type of large building. The biblical version more naturally resembles a kind of sea vessel. In the Mesopotamian tradition, a large number of survivors board the boat, whereas in the biblical story, only eight members of Noah's family survive.

The name of the mountain in Mesopotamian tradition is Mount Nisir, but Genesis recounts that the landing place of the ark was on (one of?) the *mountains* of Ararat (the Greek tradition reads differently). The bird scenes vary from tradition to tradition. The Mesopotamian version has only a series of three trials, in each of which a bird is sent out—first, a dove; second, a swallow; and third, a raven. In contrast, the Bible speaks of four trials, in which a raven is sent out initially, followed by three doves. Whereas the divine blessing refers to immortality in Mesopotamian tradition, in Genesis, God promises never again to destroy the Earth with a flood, a promise he seals with "his bow in the clouds." God also makes a covenant with Noah stipulating that man is forbidden to eat the blood of an animal and that whoever sheds the blood of another human will have his own blood shed. Furthermore, Noah only leaves the boat after he is summoned by God. The Mesopotamian hero leaves the boat of his own will.

Of course, in the biblical account, one singular deity is all-powerful throughout the entire flood narrative. At variance with this, in Mesopotamian tradition, the announcement of the flood takes place in the divine assembly—a meeting of the gods to take counsel and make decisions concerning the fate of humanity as well as the gods themselves. The Mesopotamian assembly of gods seeks to conceal its plan to destroy mankind by inundation. The biblical account does not indicate any such concealment of God's plan to flood Earth.

Consequently, recent scholars in Old Testament and Assyriological studies have concluded that there was never any sort of direct borrowing from the Babylonian flood account into the biblical account or vice versa. This has given rise to the view that two independent traditions have descended from a common original tradition. While this makes sense of both similarities and differences between the two accounts, we possess no real evidence to test the validity of this proposal. Would this common tradition have been in written or oral form? The evidence from Mesopotamia demonstrates, too, that there was certainly more than one distinct tradition of flood and creation stories. This is even evident in the differences of the flood stories between *Gilgamesh* and *Atrahasis*. One may argue further that, although there was a common conceptual framework concerning the flood, texts at our disposal indicate much diversity and diffusion of views in the ancient Near East.

How is the borrowing from one literary text into another detected and measured? The following comparative spectrum could be proposed as John Walton (2006) has done in *Ancient Near Eastern Thought and the Old Testament*, moving from differences to similarities: totally ignores and presents a different view altogether; hazy familiarity leading to caricature and ridicule; accurate knowledge resulting in rejection; disagreement resulting in polemics, debate, or contention; awareness leading to adaptation or transformation; conscious imitation or borrowing; subconscious shared heritage. Furthermore, do the documents in comparison have a distinct or like purpose, do they employ similar or distinct textual genres, and are the similarities in vocabulary—especially in the case of etymology—enough to establish a readily identifiable connection among texts separated in space and time?

Patriarchal Narratives

Delitzsch did not go into any detail about the historicity of the patriarchal narratives—the stories about Abraham, Isaac, Jacob, and Joseph, as contained in Genesis 12–50—or whether they had been influenced heavily by Babylonian culture. In the following decades, newer finds from Near Eastern sites such as Mari, Nuzi, Alalakh, and Emar gave rise to much speculation over the authenticity of the patriarchal accounts. Scholars such as John Van Seters contended that the stories of Abraham and his descendants did not reflect the time period in which they were purported to have taken place (early second millennium BCE); rather, they reflected only the views of later generations (first millennium BCE). These later narratives, in their estimation, were mere fabrications of Jewish writers to serve the purposes of their immediate social context. References to camels and the "Philistines" in these earlier episodes of Israelite history were alleged as only coming into Israel later than the time of Abraham by approximately a millennium. While the domestication of camels in Mesopotamia (the origin of Abraham's descendants) was actually quite early, all must admit that the Philistine peoples had not yet inhabited the region during the time of the patriarchs.

Early responses, however, to Van Seters were similar to those that came against Delitzsch after his initial lecture in 1902, mainly along party lines. Furthermore, no evidence surfaced to favor any one opinion. To date, there are no texts mentioning Abraham or his family by name, nor should we expect simple sheepherders to be mentioned in royal or private archives of Mesopotamia. The problem of continuity in Near Eastern culture must be acknowledged, namely, that the traditions of the Near East appear to be very durable and resistant to change. As a result, many of the customs and patterns present in the patriarchal narratives, such as nomadism, legal customs, and religious practices, could also have been at home in later periods of Israelite history. Therefore, if proof cannot be furnished of cultural features that can *only* be attributed to the time of the patriarchs, then this is a

Artist's rendering of the temple and palace complex of Ur in Abraham's time. (Balage Balogh/Art Resource, NY)

weakness inherent to a belief in the historical reliability of these portions of the Hebrew Bible. However, scholars continue to argue that they can furnish proof of cultural characteristics that cannot belong to a later time in Jewish history.

Dan Fleming (2003) has recently argued this in "History of Genesis" on the basis of evidence from the ancient site of Mari, where roughly 20,000 cuneiform documents have been unearthed. He argues for a connection between the "Benjaminites" of the Hebrew Bible and the *Binu Yamina* "Sons of the Right (that is, north)" of the Mari archives. He suggests, moreover, that these "Benjaminites" possessed ties in antiquity with the famous Mesopotamian site of Harran—also known as the place where Abraham's ancestors had settled after coming from distant Mesopotamian inland to the south. The vicinity of Harran was the birthplace of many of Abraham's later descendants. The Mari documents suggest a strong link between the site of Harran and the tribe of Benjamin. In fact the word "Hebrew" may itself be related to a term used to designate herding groups among the *Binu Yamina* (*'ibrum*). In contrast to this group from the north, there existed at Mari the *Binu Sam'al* "Sons of the Left" (that is, south). In the Hebrew Bible, there is a passage that is at least interesting in light of these parallels:

Abram said to Lot, "Let there be no strife between me and you, between your herdsmen and mine, for we are men (and?) brothers. Is not all the land before you? Separate yourself from me. If you (go to) the left (*smo'l*) I will

go to the right; if you (go to) the right (*yamin*) I will go to the left." (Gen. 13:8–9; author's translation)

Comparable to the Mari material, the biblical texts define pastoral landholdings by directional metaphorical language corresponding to "right" and "left." It may be that the Bible has preserved an ancient conception of tribal separation in the second millennium BCE. Like Fleming, other scholars believe that the customs of the patriarchs are authentic to the era and are further corroborated by finds from the ancient Near East. Furthermore, they believe much of what Van Seters has asserted to be falsifiable. For example, the few references to the Philistines in Genesis are different from references to them in later books of the Old Testament. It could be that the authors (or updaters) of Genesis employed the term Philistines to refer to groups of non-Canaanites who were immigrating into the land of Canaan from the Aegean.

Interpretation: Conservative Views

In the following days after Delitzsch delivered his lecture(s), an ensuing debate developed over the relationship of Babylon and the Bible. On the conservative front, counterarguments were limited mainly to philological problems, that is, matters of language, individual words, and smaller textual issues. However, the real issue at hand was the underlying theological debate over the Old Testament. Scholars felt that the inspired revelatory character of the Old Testament had been threatened. Against the belief that it was a supernatural document, cuneiform findings were regarded by some as having demonstrated its basic human origin. Some were of the view that, if the uniqueness of the Old Testament in relation to its surrounding environs could not be demonstrated, then a belief in its revelatory nature would be difficult to embrace. Thus with each correspondence between the biblical and Babylonian cultures, the death knell appeared to have been sounded against the uniqueness and inspiration of the Old Testament. In its most traditional form, the ultraconservative position was very plastic in its view of scripture. In reality, however, many of them were only moderately conservative, allowing for similarities between the two cultures, though not to the extent that Delitzsch had suggested.

One of the major points of contention was Delitzsch's reading of the name of Yahweh in cuneiform documents from 2500 BCE. Very few scholars, even in the early days, had accepted this reading of the text, and since then, the dating of the texts, the reading of the name of Yahweh, and the text's ties to King Hammurabi of Babylon, who was supposedly related to the Amraphel of Genesis, have all been adequately rebuffed or updated.

Conservative scholars were also able to make use of Babylonian evidence to support their views of the Bible. Since early accounts of creation and flood stories

had been recovered of Mesopotamian origin that contain several parallels in common with biblical material, the traditional (early) dating of certain portions of the Old Testament appeared to be vindicated. The Mesopotamians clearly regarded the flood as a truly historical event because they included it in their kings' lists. It was as much a defining event for them as it was for the Hebrew authors.

Interpretation: Liberal Views

On the other side of the debate was the liberal point of view, which, in the most extreme cases, had already renounced the belief in the verbal inspiration of scripture. For them, Delitzsch's "low" view of scripture was not problematic or even new for that matter. Still, many liberals were of the opinion that the Hebrew scriptures were inspired, but that the traditional view of inspiration had to be altered in view of the scientific study of scripture. This approach was championed by such scholars as Hebrew/Semitics Yale professor William Rainey Harper, who later became the founding president of the University of Chicago in 1892.

Liberals sought to protect the views of critical theology, which had accumulated after years of research and debate. Problematic to this goal was that the theory of the literary development of the Pentateuch had grown in a discussion void of nonscriptural evidence, such as the discoveries from Mesopotamia of which Delitzsch spoke. Liberals believed that the majority of the Bible, at least in its final form, had been written in the period after the Babylonian exile of the Israelites (587 BCE). Complex theories were devised to determine the sources lying behind the Old Testament. When Delitzsch claimed in *Babel and Bible* that "Assyriology has re-established the credit of *The Tradition of the Old Testament Text*, which has so long and so fiercely been assailed" (Delitzsch 1903: 166), scholars of the liberal school began to oppose him. They also quibbled about his claim to have found the name of Yahweh in Babylonian documents. In this they were very close to conservative scholars, even though their underlying motives were surely unrelated to the latter. They were prepared to agree with Delitzsch about the heavy influence of Babylonian culture on the Bible, unless it proved the ancient authenticity or coherence of narratives was thought to be incoherent (confirming the need for hypothesized sources).

Though Delitzsch was misled in many of his arguments, as it happens, the German Assyriologist had struck a nerve in Old Testament studies, bringing to the fore the deep chasm between conservative and liberal scholarship, and doing so with fresh and hard evidence from the provenance of ancient Babylon. This made the daggers of his argumentation all the more piercing. To the conservatives, he represented a danger to their idealistic views of the Bible, particularly with respect to the theological doctrine of inspiration. To the liberals, ironically, he challenged their views by claiming that the recent cuneiform findings corroborated much of biblical tradition. Hard-to-explain words, which were often

regarded as errors and emended, were verified and explained on the basis of new evidence. The Bible, to some extent, was justified by Babylon.

Confirming the Old Testament

The two opposing tendencies of liberals and conservatives can be observed, depending on whether or not new texts surfacing from the ground show a close affinity to the Bible. First, if there is believed to be a close affinity with the Bible, the texts are used by liberals or secularists to demonstrate that the Bible is not unique and therefore not divine revelation. On the flip side, the Bible is often demonstrated to be reliable in its telling of historical events by use of the same documents. Second, if portions of the biblical text are without literary parallels from the ancient Near East, the Bible, though regarded as unique, has no way of being verified scientifically with observable nonscriptural literary parallels. The belief in a unique Bible comes at the expense of having little or no way to confirm its historical reliability. In contrast, literary parallels that demonstrate the authenticity or reliability of the biblical texts come at the cost of the Bible's unique inspirational message.

More recently, scholars have argued convincingly that the Bible, in order to be inspired, must show similarities with the culture of its surrounding environs. In fact, the uniqueness of the biblical message can be observed *only* because of its resemblance to the broader culture. This view has been espoused by Peter Enns (2005) in *Inspiration and Incarnation* and Walton in *Ancient Near Eastern Thought and the Old Testament*. Enns and Walton believe that the modern reader is often blinded to the message of the Old Testament because of his or her current expectations that result from reading modern literature and history.

We cannot read the Old Testament like a modern journalistic or historical textbook. Such reading would be contrary, in many cases, to its purpose and values. Most of the biblical stories are not eyewitness accounts as is most of modern history. Therefore, both critical and confessional scholars need to rethink their views of the text. They must reckon with the limitations of science, namely, that it cannot nor should be the final arbiter between history and faith. Though science often corroborates the historical reliability of the Bible, as Kenneth Kitchen (2003) has fully demonstrated in *On the Reliability of the Old Testament*, most often we simply do not have the kind of direct evidence necessary to validate scientifically many of the claims of the Bible. Nor should we expect such evidence, in many cases, ever to come to light. Furthermore, since many of the modern concepts attached to scientific empiricism were unfamiliar to the ancient world, we should not expect ancient readers to write in accordance with our modern principles of historical composition. When a deity is said to act in historical events, such claims cannot be proved or said to be erroneous on the merits of scientific study. Miracles and the actions of the divine belong to the realm of faith, not to the realm of empiricism.

Since in almost every case the Bible has demonstrable similarities *and* differences with the ancient world, it can be said that the Bible is authentic *and* unique. One is able to believe in the divine inspiration of its message and still be intellectually honest with the results of modern science. Inasmuch as the Bible is analogous to the prevailing culture, so too its differences relate to how God *transformed* the stories of the ancient culture so that he became the focal point of their message. In contrast to a dysfunctional family of divinities, as we observe, for instance, in *Enuma Elish, Gilgamesh*, and *Atrahasis*, the God of the Old Testament is amazingly unique and just. His actions are regarded as ethically upright and perfect. Being divinely inspired, the text of the Old Testament was effective and sufficient to get across its message to Israelites and Jews of its time. It also was influential in the developing concepts and beliefs relayed in later Judaism, Christianity, and Islam, never mind its place of honor in Western civilization.

Conclusion

In the final analysis, who can say whether the events of creation and the great flood did not actually happen as they were recorded in the book of Genesis? Can science or modern literary theory? For the Mesopotamians as well as the biblical writers, they were facts of history. Once the Bible is viewed on the basis of its own cultural assumptions, it may be regarded as self-confirming.

References and Further Reading

Chavalas, M. W., and K. Lawson Younger. *Mesopotamia and the Bible*. London: Sheffield Academic Press, 2002.

Delitzsch, F. *Babel and Bible*, Vols. I and II with notes. Edinburgh: Neil, 1903.

Enns, P. *Inspiration and Incarnation: Evangelicals and the Problem of the Old Testament*. Grand Rapids, MI: Baker, 2005.

Finkelstein, J. J. "Bible and Babel: A Comparative Study of the Hebrew and Babylonian Religious Spirit." *Commentary* 26 (1958): 431–44.

Fleming, D. "History in Genesis." *Westminster Theological Journal* 65, 2 (2003): 251–62.

Heidel, A. *The Babylonian Genesis: The Story of Creation*. Chicago: University of Chicago Press, 1942.

Heidel, A. *The Gilgamesh Epic and Old Testament Parallels*. Chicago: University of Chicago Press, 1946.

Hess, R. S., and D. T. Tsumura. *I Studied Inscriptions from before the Flood: Ancient Near Eastern, Literary, and Linguistic Approaches to Genesis 1–11*. Winona Lake, IN: Eisenbrauns, 1994.

Kitchen, K. A. *On the Reliability of the Old Testament*. Grand Rapids, MI: Eerdmans, 2003.

Larson, M. T. "The 'Babel/Bible' Controversy and Its Aftermath." In J. M. Sasson, ed., *Civilization of the Ancient Near East* (pp. 95–106). Peabody, MA: Hendrickson, 1995.

Van Seters, John. *Abraham in History and Tradition*. New Haven, CT: Yale University Press, 1975.

Walton, J. H. *Ancient Near Eastern Thought and the Old Testament: Introducing the Conceptual World of the Hebrew Bible*. Grand Rapids, MI: Baker, 2006.

Wind, J. P. *The Bible and the University: The Messianic Vision of William Rainey Harper*. Atlanta: Society of Biblical Literature, 1987.

CON

Many of those who were first to travel to the Near East from the West to study the archaeological remains of ancient cultures did so with the Bible in mind. Frequently these early archaeologists were members of the clergy. They and some lay archaeologists sought to bolster the Bible's claim to historical accuracy by unearthing evidence supporting it. Even early scholarly archaeologists, such as William Foxwell Albright, believed that archaeological evidence could offer biblical proof. The beginnings of what was called biblical archaeology saw a search for sites named in the Hebrew Old Testament and evidence for the historicity of individual people named in the texts. While it has been decades since these early works were published, these ideas remain a legacy to modern archaeologists. It should come as no surprise that archaeology of the Near East still attracts those who wish to prove the Bible is true and supports Judeo-Christian beliefs. Some individuals—religious, scholarly, and lay alike—have interpreted their occasional "finds" as evidence for a broader confirmation of the biblical historicity. Proposed sources had come from throughout the Near East, including sites relating to the neo-Assyrian and neo-Babylonian empires.

In the early to middle first millennium BCE, the ancient Near East saw the beginning of the rise of empirical powers. Between 722 and 598 BCE, the neo-Assyrian Empire held Israel in subjugation and the southern nation of Judah as a vassal. The neo-Assyrian conquest spread throughout most of the Near East, placing this region in a position of utmost political power. But in the late seventh century BCE, the neo-Babylonians rose against the Assyrians. Beginning in 598 the neo-Babylonian Empire began to exert power over Judah, and by 587 it held control over the southern kingdom. However, the neo-Babylonians remained in power only until 539, when the Persians under Cyrus came to power. One method used by the neo-Assyrian Empire and the neo-Babylonian Empire to maintain their

hegemony over a broad portion of the ancient Near East was the forceful exile of individuals from their homes to be resettled in different parts of the empires. The Assyrians used this tactic in their conquest of Israel, but left the Judeans, instead exacting tribute from them. When the neo-Babylonian Empire was faced with rebellions, it forced some of the people in Judah into exile in Babylon.

There was clearly a prolonged period of cultural and political contact between Judah and Israel and both the neo-Assyrian and neo-Babylonian empires. Because of the large period of time involved in contact and exile, it should be expected that some evidence of the events and people involved should appear in the neo-Assyrian and neo-Babylonian sites. Archaeologists and epigraphers have found such evidence in many different sites. However, there is a large difference between proving a single event or the historicity of a biblical individual and confirming the Bible's narratives as a whole. Based on archaeological evidence, limitations, and theory, any pronouncement that excavations in neo-Assyrian or neo-Babylonian sites have proved the Bible as historical must be seen as oversimplification at best and at worst a false interpretation of what has been found.

Use of Assyrian Written Record

The neo-Assyrian Empire and the neo-Babylonian Empire both represented complex societies, which required large amounts of recordkeeping. Because of this, writing was an important part of both cultures and turns up frequently in the archaeological record. Included among such remains is an inscription found at the Assyrian site of Kurkh on a monolith. This inscription mentions Ahab of the land of Israel. This appears to match with a king known in the Hebrew Old Testament and seen in the book of I Kings 21–22. Thus the monolith inscriptions provide extra-biblical evidence for the existence of a biblical character. Many attempts have been made to prove biblical historicity based on such evidence. Aspects of the Kurkh monolith make it difficult to believe that the mention of Ahab of Israel is capable of acting as a support of the Bible as a larger whole.

The monolith found at Kurkh possesses a number of specific contradictions to the biblical text, with which its mention of Ahab of Israel appears to mesh. The monolith mentions Ahab as part of a combined force of kings who fought against the Assyrian king Shalmaneser III at Qarqar. However, the Bible does not mention that such a battle occurred, despite its apparent importance and the large scale it seems to have had, either 11 or 12 kings fighting together against Shalmaneser III. Additionally the forces of King Ahab are enumerated in the Kurkh inscription as being composed of 2,000 chariots and 10,000 soldiers. This is contradictory to the strength of the Israelite forces under Ahab as noted in the book of I Kings 20:15 and 20:23–25. The numbers not only disagree,

with the Israelites being noted as having a combined force of 7,000 men in I Kings 20, but they also do so on a large scale.

Clearly the weak Israelite army noted in the book of I Kings does not fit with the strong force represented in the Kurkh monolith. Two interpretations of this issue have been presented. First, it has been theorized that the numbers in the Kurkh inscription represent a coalition of different nations' armies together under the control of Ahab. This too is not known from the Bible. Alternatively, it has been proposed that I Kings 20 and 22, which describe the weak army of Israel, may in fact include an editing error. These verses may have originally been written about King Jehu's reign as seen in II Kings 13. This may have been possible, as the verses in I Kings predominantly use the term "king" instead of mentioning a specific ruler by name. During this later period, the Israelite army would have been weakened and more along the lines what was presented in the book of I Kings.

The inscription on the Kurkh monolith names a specific biblical figure, Ahab of Israel. However, the extra-biblical support of this individual does not support the actual text of the biblical narrative. Instead, it offers reports of a large battle that is absent from the biblical histories and presents hosts that are incongruent with those represented in the texts. The mention of the Battle of Qarqar, although absent from the Hebrew Old Testament, does not disprove its narrative nor does the alternative to the size of the armies in the Bible. However, the questions raised by these discrepancies require that the history of the Bible be noted as incomplete and possibly that redactors may have altered its historicity.

Use of Babylonian Written Record

Much like the Assyrians, the Babylonians left written records that dealt with figures found in the Hebrew Old Testament. Of these figures, one that is easily found is Nebuchadnezzar II, the ruler of the neo-Babylonian Empire during its initial conquest of Judah. There are many formulaic inscriptions left by Nebuchadnezzar. These inscriptions, and those of other Babylonian rulers, often deal not with their conquests and campaigns, but with portraying them as guardians of the temples and tireless patrons of their cults. This is contrary to the Old Testament's portrayal of Nebuchadnezzar. The biblical text is much more concerned with his effect on Judah, the conquest, and later exile of its people.

Despite the large differences in focus between the majority of Nebuchadnezzar's inscriptions and the writings of the Old Testament authors, there do exist some examples of Babylonian texts that offer support for the biblical history. One such text, BM 21946 of the Babylonian Chronicles, is a cuneiform inscription that tells the history of the early years of Nebuchadnezzar's rule. Included in this history is his siege of Judah in 597 BCE. This event is recorded in II Kings 24:8–17 and Jeremiah 52:28–30. This seems to be a clear

occurrence of extra-biblical support for an event in the biblical text. In this case of Jeremiah 52:28–30, it is such a validation; but the text of II Kings 24 presents an issue. While both the Babylonian chronicle and the book of Jeremiah note that the siege occurred in the eighth year of King Nebuchadnezzar II of Babylon's reign, II Kings 24 places the battle in the seventh year of Nebuchadnezzar II's rule over Babylon. While this may seem a minute point in the larger concordance of the siege itself, it still presents a sticking point in attempting to confirm the Old Testament. Clearly, if the text of II Kings 24 not only disagrees with extra-biblical texts but also with the biblical text in Jeremiah, there is an issue with any attempt to present this text as historically accurate or to say that the Babylonian chronicles actually confirm the Old Testament. It appears more likely that the cuneiform tablet can be used to verify a single event also mentioned in the Bible, but not the Bible itself.

While Nebuchadnezzar is mentioned in Babylonian material remains and the biblical narratives, this does not mean that every mention of Nebuchadnezzar found in the Old Testament shares historical authenticity with him. One need only read the book of Daniel to find evidence of this. Daniel 2–4 deals with a group of Judean exiles in Babylon and the interaction of one of them, Daniel, with King Nebuchadnezzar. The various interactions between King Nebuchadnezzar and the captives cannot be validated as historical through extra-biblical sources, as can the 597 BCE siege of Judah. The book of Daniel is a work compiled long after the neo-Babylonian conquest, and any historical portion of it is written in the third century BCE at the earliest. Additionally it does not appear to have been written as a history but rather as a theological piece, presenting a specific understanding of the nature of YHWH (Yahweh; God). For instance, in Daniel 4, Nebuchadnezzar is brought low by YHWH and forced to live as an animal for seven years. Following this period, Nebuchadnezzar is said, in Daniel 4:31, to have praised God and followed his will. There is certainly no evidence that Nebuchadnezzar ever worshiped or praised the Judean god. Neo-Babylonian inscriptions describing the works of Nebuchadnezzar clearly show him offering praise to the Babylonian deity Marduk. This is clear counterevidence to Daniel 4, and it illustrates the inability to confirm the Bible through a historic individual's name.

Difficulty of Matching Archaeological and Written Records

The complex process and long period of time that went into the creation of the books of the modern Hebrew Old Testament mean that some such historical stories, such as those in the initial chapters of Daniel, cannot be well substantiated by archaeological evidence. In fact, it appears some later writings concerning the exilic period clearly go against extra-biblical sources, such as those that illustrate Nebuchadnezzar II's worship of Marduk. The neo-Babylonian Empire

A Kassite relief sculpture of Marduk, the main deity of Babylon, with a dragon. (Bettmann/ Corbis)

may have left some evidence of characters or events in the Bible, but this has not confirmed the Bible; it has only offered further critiques of its accuracies and inaccuracies.

The mention of an act, person, or place in the chronicles of the neo-Assyrian and neo-Babylonian empires does not confirm every story of them in the Hebrew Old Testament, but what happens when there is no evidence of an event provided in the chronicles of these two powers? Such an instance occurs concerning the destruction of the first Jerusalem temple in 586 BCE. The Babylonian chronicles are fragmentary, and gaps appear in the texts concerning 594 to 558 BCE. Thus, there is no supporting history for the destruction of the temple, one of the most important events of the exilic period to the biblical authors. This event is mentioned in II Kings 25:9–17, II Chronicles 36:19, and Isaiah 64:11. Evidence of the destructive campaign of 586 BCE has been found at sites throughout Judah and has been dated to the proper time. However, the neo-Babylonians did not leave collaboration to this event in their own official documentation, or at least those discovered so far. There appears to be no reason to doubt that the temple was destroyed in 586 BCE under the orders of Nebuchadnezzar II, but it would not be proper to say that finds in Babylonian sites have confirmed this important historical event on which the authenticity of many biblical passages depend. It is also important to note that the Babylonian chronicles have many other gaps. The years 556, 552 to 550, and 544 to 540 BCE are all unrepresented in the Babylonian chronicles, which thus cannot be offered in confirmation of any events in these years.

Much of the evidence that has been assessed to this point has been textual, cuneiform pieces that have been found in the ancient ruins of the neo-Assyrian and neo-Babylonian empires. These texts can serve to tell much about the people, events, and places that existed in the influence of these two political powers and have given much aid to the assessment of the Bible's claim to historicity. Texts offer many important advantages over other material remains that archaeologists

unearth. They tell a story in words, which requires less interpretation by the archaeologist. For instance, seals can name a bearer, telling who used it, possibly for what reason, or his position. Because of this, it is much more likely to find support for something that is written in the Old Testament in an Assyrian or a Babylonian text than in other nontextual remains from such sites.

The role that texts can play in offering support for the events portrayed in the Hebrew Old Testaments is great, but various issues with the texts of the neo-Assyrian and neo-Babylonian authors temper it. Perhaps the greatest flaw in relying on texts is gaps, such as those found in the Babylonian chronicles, which lack documentation over a broad range of years that included important events. A similar gap appears in the royal inscriptions of Sargon II, king of the neo-Assyrian Empire. His inscriptions go without mentioning the deportation of some Israelites from Samaria that is seen in II Kings 17:6 as well as other Assyrian sources. The addition of other Assyrian sources gives support for this event, but the presence of multiple sources cannot always be relied on as it can in this situation. Additionally texts that do survive to be found are often in poor physical condition, with breaks in the text that must either be extrapolated or lost. Within these spaces, important names or similar data can disappear from the historical record. However, these gaps in the presence of writing and in the physical texts themselves do not disprove the Bible. But they rob scholars of the ability to confirm it as well.

Perspective and Hyperbole of Text Writers

While the absence of texts offers no support to the Bible, even the study of extant and intact texts can itself provide issues in looking for support for the biblical text. The writings found in neo-Assyrian and neo-Babylonian contexts represent the beliefs and words of a specific and select group of people. Scribes associated with the government and temple would have been responsible for much of the written remains found by archaeologists. These texts would have represented a limited frame of mind, that of the scribes and those on whose behalf they were writing. Textual archaeological finds possess an intent that other material remains do not. For instance, the biblical authors had a specific intent when they wrote about Nebuchadnezzar II. He had destroyed the most important piece of their cult and brought their divinely appointed ruler into exile. Resulting from this is a negative portrayal of Nebuchadnezzar and Babylon as whole within the exilic and postexilic writings in the Hebrew Old Testament.

A contrary writing comes from the Wadi-Brisa Babylonian inscription. In this, Nebuchadnezzar II is portrayed as a ruler who protects the subjects of his nation, leads them home from far-off lands, constructs cities and roads, and does it all for the glory of his god, Marduk. Much of this runs contrary to the biblical portrayal of the neo-Babylonian ruler, but especially the idea that he

led scattered people home to their own settlements. This is in direct opposition to weight given to Nebuchadnezzar's removal of the Judean upper class. While this incongruity does not falsify the biblical narrative as a whole, or even the events of the exile or the conquest of Judah by Nebuchadnezzar II, it does make evident the agency of the biblical authors. The figure of Nebuchadnezzar II in the Hebrew Old Testament cannot be seen as truly historic, but as the invention of authorial agency and intent, a character invented to serve a role in a broader historic theology.

In addition to issues of portrayal of historical figures in texts, the nature of recorded events can be problematic. Many of the existing royal inscriptions from the neo-Assyrian Empire offer evidence of this. These writings, which tell of the deeds of the kings, are often full of hyperbole and exaggeration when compared to the more administrative texts from the same period. This creates an issue of credibility when looking at these texts for support of the biblical narratives. Even when events can be matched between the exaggerated texts and the biblical narrative, the specific details can be problematic because of the neo-Assyrian sources' hyperbole. These problems with the royal inscriptions can only be partially overcome by focusing on the administrative texts found in the form of epistles that are usually more factual. However, their nature makes it difficult to apply them to historic events that the Bible may mention. They are often quite terse, dealing with details and short pieces of information. Additionally, the use of such writings did not necessitate that the date be recorded on them. Thus it is not only difficult to find events recorded on these pieces that can be compared to other historical narratives, but also to accurately date and provide a context for each individual text. This lack of context makes it difficult to form conclusions of the historicity of the epistles—and subsequently difficult to apply them correctly. Although the textual remains of the neo-Assyrian and neo-Babylonian empires offer one of the most fertile areas in which to find evidence concerning the Bible's historicity, these sources also suffer from issues that make it difficult to fully apply them to the biblical texts.

Lack of Archaeological Corroboration

While many material remains and some textual remains may be problematic when looking for clear understandings, some archaeological evidence does present relatively clear evidence of historical occurrences. A good example of this is the Lachish relief found in Nimrud. The relief was spread around the walls of a room in the southwest palace. It depicts an Assyrian attack on the Judean city of Lachish. The relief is usually tied to the campaign of 701 BCE, although the inscription on the Prism of Sennacherib, which records this campaign, does not record this event. The archaeological evidence from Lachish is also questionable. There are two evident layers that represent destructive attacks on the city.

Some scholars attribute both layers of destruction to the armies of Nebuchadnezzar II, one from 597 BCE and the other from 587 BCE. Other researchers follow the evidence of the Lachish relief to date the earlier destruction layer to Sennacherib in 701 BCE and the second to Nebuchadnezzar's conquests. Unfortunately the Lachish inscription does not present proof as to whether or not Sennacherib destroyed the city of Lachish in 701. While it presents a relatively clear support for an Assyrian siege at Lachish during Sennacherib's reign, the relief found at Nimrud does not offer clues into the interpretation of other archaeological remains. This is a clear example of the issues present in nontextual archaeological remains.

While the Lachish relief may offer only a brief glimpse of what went on in the city when the Assyrians were present, it is still relatively strong support for the presence of the Assyrians and an associated siege. The omission of the events from Sennacherib's other inscriptions does not mean that the event did not occur, but that it was just likely not recorded there. Lachish was one of the largest and most important cities in the Judean kingdom. Any battle at such a center would have been of great political and economic importance, especially since Lachish's position served to assist in the fortification of Jerusalem. Thus, if the Assyrians had laid siege to Lachish, one would expect that it might be mentioned within the Bible's historic narrative. Lachish is mentioned numerous times in conjunction with the Assyrian army's presence in Judah. The book of II Kings 18:14, 18:17, and 19:8 and the book of II Chronicles 32:9 tell of the Assyrians sending messages from Lachish or moving their forces from the city, but neither book clearly states that the Assyrians had conquered, let alone destroyed, Lachish.

There are many possibilities as to why the actual siege is not mentioned in the Old Testament. It is possible the writings were meant to imply that Assyria had been in control of Lachish, as they are noted as being at the city during a period of war with Judah. Alternatively it may have been completely overlooked by the biblical author for some reason that has not been discovered. The alternative that the Lachish relief is itself problematic and not historical has not met with acceptance, as there appears no reason to doubt it, and it does fit an interpretation of the associated biblical passages. It is clear that the Lachish relief can only be seen to confirm the Bible in that it shows Assyrians at Lachish. The siege, which it illustrates, either contradicts the Bible's lack of such a story or expands on the modern understanding of the text, showing that it presents a history that is missing important events such as the siege of one of the largest Judean cities.

The manifold examples of archaeological remains from neo-Assyrian and neo-Babylonian sites that pertain to the historicity of the Hebrew Old Testament's narratives provide a varied amount of success. None of them offer the ability to say that the entire Bible is supported or even that the events with which they deal are entirely confirmed. In this entire discussion, the issue of role of archaeology has yet to be raised. Many of the earliest archaeologists in

the Near East went with the mission of proving the Bible as true and confirming its historical narratives. It has become incumbent on modern scholarly archaeologists to define their role in such contention and address the role of their field in the confirmation of the historicity of the Hebrew Old Testament's stories. Beyond this issue, it is necessary to lay out the possibility of such confirmation within the broader understanding of archaeological methods and theory.

Confirmation Issue in Decline

The issue of confirming the Bible as a historically accurate document has been in decline within the recent work of Near Eastern archaeologists and other scholars. Focus has instead been on using archaeological remains to better understand life in the ancient Near East and to propose accurate histories for this region. While these histories are free to take into account the biblical text, they do not give it special weight when compared to contrasting histories and evidence present in the material remains that are unearthed. There are many reasons that such methods and goals have been adopted. Among these is recognition of the nature of the biblical text. Written over the course of many centuries, by many different authors, the Bible has been consistently composed and revised not only to present historic narratives but also to illustrate theological ideas. This conclusion has been reached through the analysis of the Old Testament using text critical methodology and in the light of continuing archaeological work in ancient Near Eastern sites.

A good example of this comes from excavation at the biblical city of Jericho. The early excavation at Jericho by John Garstang unearthed walls that had apparently collapsed catastrophically. Garstang dated the destruction to the 15th century BCE, contemporary to the period of settlement of Canaan by the future Israelites as seen in Joshua 6. This was seen as a strong proof of the biblical

---

The Historicity of David

There are undeniably some discrepancies in the Bible's accounts of King David (he meets Saul for the first time on two different occasions, and another man, Elhanan, is referred to as Goliath's killer in II Samuel 21:19), but even aside from them there is the question of whether or not he was a real historical figure around whom legends grew. There are two clear references to a house of David on inscriptions uncovered in Israel and Jordan, but this establishes only that the monarchy of Israel had claimed descent from David in ancient times, to which the simple existence of the Old Testament already attests. Other than the name and timing, there is nothing to connect this house of David with a king who unified Israel and Judah, defeated a giant, wrote psalms, won the battles associated with David, and so forth.

historicity and a success of biblical archaeology. However, in the 1950s Kathleen M. Kenyon reassessed the site and found that the dating was incorrect, pointing instead to the 16th century, which was too early for the biblical narrative to be supported. The inability of archaeologists to find walls at the Jericho site that can be dated to the time of Joshua is a clear illustration of both the problem of biblical historicity and the error of basing one's work on searching for such.

The active search for archaeological support for biblical historicity has not only waned in response to further understanding of the nature of the Bible, but also in response to the ever-evolving science of archaeological methods. The larger discipline of archaeology has become continually more scientific. One aspect of this development is the methodological belief that evidence should speak for itself. Searching for specific evidence of the Bible's historicity does not allow for this, because it posits the same importance to the biblical narrative and searches for a specific conclusion, instead of arriving at whatever ends the evidence leads to. Garstang's work has often been cited as an example of this problem.

Conclusion

Archaeologists have found many pieces of evidence that have offered some amount of confirmation of the Old Testament. However, this confirmation has been of limited importance and scope. Textual and physical remains have given support to many biblical figures and events, but at the same time they have continually introduced questions concerning the historicity of the biblical narrative. How could Ahab have sent an army to Qarqar that was so much larger than that presented in the book of I Kings? How could Nebuchadnezzar II have worshiped YHWH while he is shown in his own writings working for the glory of Marduk? How could the biblical authors overlook the siege of an important city like Lachish but specifically note time and again that the Assyrians were at Lachish? These questions and many others make evident the simple fact that not only have extra-biblical sources failed to confirm the Old Testament narrative as a whole, but also that such an event is unlikely. This is because of the nature of the Bible. It may be partly a historic document, but at the same time it contains a theological aspect, which has at times, such as the book of Daniel, superseded the historical. Other issues may arise because of the long process of writing, rewriting, and redaction that has gone into the composition of the biblical text. Such changes may explain why Ahab's small biblical army was able to appear in such force at Qarqar. Perhaps it is best to conclude by presenting the idea that the historicity of the Bible is not supported by archaeology, but by faith, as Shalom M. Paul and William Dever (1974) posit in *Biblical Archaeology*. For to believe that the entire Bible can be confirmed as a single historical document requires that one look at it not in the light of textual criticism and archaeology but in the context of religious understanding.

References and Further Reading

Albertz, Rainer. *A History of Israelite Religion in the Old Testament Period*. First American ed. Old Testament Library. Louisville, KY: Westminster/ John Knox Press, 1994.

Albertz, Rainer. *Israel in Exile: The History and Literature of the Sixth Century B.C.E.* Boston: Brill, 2004.

Fagan, Garrett G., ed. *Archaeological Fantasies*. New York: Routledge, 2006.

Kelle, Brad E. "What's in a Name? Neo-Assyrian Designations for the Northern Kingdom and Their Implications for Israelite History and Biblical Interpretation." *Journal of Biblical Literature* 121, 4 (2002): 639–66.

Kenyon, Kathleen Mary, and P. R. S. Moorey. *The Bible and Recent Archaeology*. Rev. ed. London: British Museum Publications, 1987.

Oded, Bustenay. *Mass Deportations and Deportees in the Neo-Assyrian Empire*. Wiesbaden, Germany: Reichert, 1979.

Paul, Shalom M., and William G. Dever. *Biblical Archaeology*. New York: Quadrangle/New York Times Book, 1974.

Pritchard, James Bennett. *Ancient Near Eastern Texts Relating to the Old Testament*, 2nd ed. Princeton, NJ: Princeton University Press, 1955.

Pritchard, James Bennett, and William Foxwell Albright. *The Ancient Near East*. Princeton, NJ: Princeton University Press, 1958.

Sack, Ronald H. "Nebuchadnezzar and Nabonidus in Folklore and History." *Mesopotamia* 17 (1982): 67–131.

Ussushkin, David. "The 'Lachish Reliefs' and the City of Lachish." *Israel Exploration Journal* 30, 3–4 (1980): 174–95.

Index

2000 B.C.E. TO 1000 C.E.

POPULAR CONTROVERSIES
in
WORLD HISTORY

STEVEN L. DANVER, EDITOR

Popular Controversies in World History

INVESTIGATING HISTORY'S INTRIGUING QUESTIONS

Volume Two
The Ancient World to the Early Middle Ages

Steven L. Danver, Editor

ABC-CLIO

I

The Ark of the Covenant is in Axum, Ethiopia.
(This idea has led to many investigations, publications, and media reports.)

PRO   Talaat Shehata
CON   Thaddeus Nelson

PRO

To better understand this argument—that the Ark of the Covenant is not lost, but that it is actually safe and in the town of Axum in Ethiopia—one needs to comprehend both the biblical and nonbiblical history of the ark. The reader will also need to gain more insight as to how throughout history different groups in many different nations have attempted to seize and, often for their own self-interest, coopted and then manipulated the concept of being God's *chosen* people. It's within that multifaceted paradigm that by the end of this section, the reader will decide for her- or himself as to the validity of the concept and the *actual* reality of a chosen people and which group, if any, should be accorded that title. But, as shall be explained, the reader will discover that the Ark of the Covenant is at the *core* of any such designation, and its possession has led some in the Ethiopian Coptic church to make such a claim.

In Exodus 25:1–2, the Old Testament states that "The LORD said to Moses, 'Speak to the people of Israel that they take for me an offering; from every man whose heart makes him willing you shall receive the offering for me.'" Then, it continues in Exodus 25:8, that the people of Israel should "make me a sanctuary, that I may *dwell* in their midst." It further stipulated in Exodus 25:9, that "According to all that I show you concerning the pattern of the tabernacle, and of all its furniture, so you shall make it." In Exodus 25:10–22, God elaborates on the primary constructs of the ark. It would be made from "acacia wood; two cubits and a half shall be its length, a cubit and a half its breadth, and a cubit and a half its height." Then, God tells Moses:

> "And thou shalt overlay it with pure gold, within and without shalt thou overlay it, and shalt make upon it a crown of gold round about. And thou shalt cast four rings of gold for it, and put [them] in the four corners thereof; and two rings [shall be] in the one side of it, and two rings in the other side of it. And thou shalt make staves [of] shittim wood, and overlay them with gold.

I

And thou shalt put the staves into the rings by the sides of the ark, that the ark may be borne with them. The staves shall be in the rings of the ark: they shall not be taken from it. And thou shalt put into the ark the testimony which I shall give thee. And thou shalt make a mercy seat [of] pure gold: two cubits and a half [shall be] the length thereof, and a cubit and a half the breadth thereof. And thou shalt make two cherubims [of] gold, [of] beaten work shalt thou make them, in the two ends of the mercy seat. And make one cherub on the one end, and the other cherub on the other end: [even] of the mercy seat shall ye make the cherubims on the two ends thereof. And the cherubims shall stretch forth [their] wings on high, covering the mercy seat with their wings, and their faces [shall look] one to another; toward the mercy seat shall the faces of the cherubims be. And thou shalt put the mercy seat above upon the ark; and in the ark thou shalt put the testimony that I shall give thee. And there I will meet with thee, and I will commune with thee from above the mercy seat, from between the two cherubims which [are] upon the ark of the testimony, of all [things] which I will give thee in commandment unto the children of Israel." (Exodus 25:11-22, King James Version)

   Finally, in Exodus 29:42–45 (King James Version), the Lord states:

"[This shall be] a continual burnt offering throughout your generations [at] the door of the tabernacle of the congregation before the LORD: where I will meet you, to speak there unto thee. And there I will meet with the children of Israel, and [the tabernacle] shall be *sanctified* by my glory. And I will sanctify the tabernacle of the congregation, and the altar: I will sanctify also both Aaron and his sons, to minister to me in the priest's office. And I will *dwell* among the children of Israel, and will be their God. And they shall know that I [am] the LORD their God, that brought them forth out of the land of Egypt, that I may *dwell* among them: I [am] the LORD their God." (emphasis added)

With that very clear knowledge in mind, the Israelites would consistently be in one in a state of endless *Shekhinah*. *Shekhinah* in Hebrew, as in Arabic with the word *S'akina*, is found in the Quran in numerous verses; the word means a state of total peace and tranquillity. In this case, this absolute reassurance and calming feeling was bestowed upon the people of Israel as a direct consequence of their individual and collective *knowing* that God was *always* with them.

   Then, in Exodus 26:31–33, a precise description is given as to how a veil would separate the ark from the priests and congregation. Even Aaron and his lineage, who were designated by God as the future priestly class, were not allowed to enter the resting place of the ark too often. They were and are expected to undertake preestablished rituals *prior* to entering the area within which the ark rested. The ark was viewed by all to be resting in an area that would become known as the Holy of Holies. That area was held in *total* piety and reverence.

With that said and done, it's quite clear that not much more was left to the imagination, let alone to be debated, within the ranks of the faithful. To all Jews, God is established, beyond any shade of doubt, to not only be the Almighty Creator of life and the universe but also to be a dynamic and active deity *within* past history and *within* its unfolding present and future. God has constantly projected and established his concern for the people who *first* chose him by declaring to them that they are his chosen ones. He would judge them, as he would other non-Jews. God would also save those he deems worth saving, as he would other non-Jews. Throughout history, God would reveal himself in multifaceted ways, either through his prophets, the historical events, the laws, or the priests. Practicing and believing Jews were of the strong conviction that God's *divine* actions have, and will continue to the end of time, to penetrate and permeate the unfolding contours of history at his own choosing. All Jews needed, and will forever need are a strong belief in God's potential and a readiness to *act* and *interact* with God's people, when he sees fit that events dictate that he would do so.

According to numerous sources, the ark contained the original Ten Commandments, Aaron's rod, and a small pot filled with manna. Manna in Exodus 16:14, is described as available in the early morning, after the dew had evaporated. It was a staple mode of daily consumption by the Israelites, which especially sustained them while they were wandering for those 40 years with Moses on their journey to the Promised Land. Similar to coriander seed in texture, it was white in color. It resembled the early morning frost on green lawns. Often, besides having it ground, it was baked by the Israelites into little cake-like or small flat doughnut (without the holes) shaped pieces. They had the taste of cake baked in oil. It was also claimed in the Old Testament that manna, especially in its raw form, tasted like a honey wafer. Gathering the manna each day, as they had been instructed by Moses, the Israelites were expected to consume only that amount for their daily sustenance. The primary reason for that directive was that it was usually extremely hard to store any of it overnight. The manna would usually decompose and begin to excrete a putrid smell. It was only on the sabbath that they were permitted to collect double the amount that they usually had grown accustomed to gathering during the week. The Israelites decided that if they doubled-up on their usual daily intake of the manna immediately prior to the sabbath, then there wouldn't be enough hours in a day for it to go completely to waste. According to this thinking, it was better to lose half the manna than all of it, so that their collective daily needs on the next day, immediately following the sabbath, could be met. It seemed to work for them. That overwhelming gratitude by the people of Israel for manna's life-sustaining qualities during their days of hardship, earned it a very special *honored* presence, next to the Decalogue stones (Ten Commandments) and Aaron's rod. It could also be perceived as the Lord's way of getting the Israelites to honor and demonstrate extreme appreciation for nature's life-sustaining *gifts*.

Historicity and the Bible

Academic interpretations of the Hebrew scriptures vary and consider the evidence in light of hypotheses about biblical sources and the relationship between biblical and outside data, a relationship that is considered on a case-by-case basis. In the case of the Ark of the Covenant, the biblical text is so specific in its descriptions that it seems almost a certainty that it was written about an actual object still known to the Hebrew people—the specificity would seem out of place and unnecessary otherwise. But the Exodus during which the Ark was meant to have been created is another matter. There is no evidence of the 600,000 people described in the Bible as having left Egypt for Canaan. Even one-tenth that number would be expected to leave considerable archaeological evidence—litter, in essence. It is entirely possible that the exodus occurred with a much smaller group, and that this group built the ark. There is also a theory that Canaan's conversion to the Hebrew religion took place not because of invasion from an outside force but because of an internal revolution, which makes it much less clear when and where the ark would have been built.

Many sections of the Ten Commandments were already existent in the laws of numerous ancient civilizations. In ancient Egypt's religious texts, the laws and prohibitions against murder, theft, and crude injustice were very well evident in the consciousness and practices of Egyptians as they entered a temple or shrine that was dedicated to their god Osiris. According to the Egyptians, he was the god that would stand judgment on them upon their death. But, most biblical and Quranic scholars are of the strong conviction that the Ten Commandments went far beyond all moral codes established by either the Egyptians or any of the other ancient civilizations. The Jewish faith represented a clean break with any of them and was explicit in its monotheism. The unsurpassed majesty and innate *goodness* of God was present for all to witness. The hidden and most intimate thoughts and desires of the human mind and heart were, for the first time, being held accountable to the highest moral laws. It was a most sacred and awesome responsibility for all who wished to believe in an incontestable Supreme Being. The Israelites clearly had their work cut out for them.

With the strong feeling of there being a *choosing*, and therefore having become a *chosen* people of God, the Israelites *knew*, as it was stipulated in Deuteronomy 7:6–11, that "the Lord [their God was] God, the steadfast God, with those who love Him and keep His commandments He keeps covenant and faith for a thousand generations. . . . Therefore, [they needed to] observe conscientiously the Instruction—the laws and the rules—with which [he charged them]." In that context, the sanctuary, which had become known as the Ark of the Covenant, also became a moving one. In the exodus out of Egypt, the ark and its bearers, who were priests, preceded the Israelites out into the Sinai Desert and

across the Jordan River. The Bible claims that with the parting of the Red Sea by Moses, with the intervention of God, the priests and the Israelites were able to cross over on dry land, until their feet touched the banks and the sea folded back onto itself. The same dry conditions, as stated in Joshua 3:15–17 to 4:10–11, 18, occurred when the priests carrying the ark, followed by the Israelites, crossed the Jordan River. The ark was present at the capture of Jericho. Prior to the fall of Jericho, and in a daily ritual, seven priests carried it, also carrying seven rams' horns as trumpets. Joshua was known to have isolated himself with the ark, to "consult it." The ark occupied a prominent location in the center of activity of the Israelites, when Joshua read to them the law. The ark was also "consulted" after the defeat in the battle against Benjamin at Gibeah.

It's within this context that one begins to get a clearer idea of how the ark, hundreds of years later, might have found its way to Axum in Ethiopia. But, first, it would have to be stolen, lost, and subjected to the possible threat of destruction or annihilation before any ruler of Israel would even entertain the thought that he and his people could temporarily part with it. In the Bible, it is stipulated that during Samuel's rule, the ark was lost to the Philistines, after Israel had withstood two separate defeats. The first defeat was at Ebenezer. Samuel and the Israelites thought very long and hard as to how they could restore their status, honor, and receding fortunes. They decided to have the ark hauled from the temple in Shiloh, where the Ark was resting, and place it before the troops, when they fought the Philistines a second time. They lost again. Only this time, the Philistines, to spite the Israelites for their demonstrated audacity and willingness to face them again in battle so soon following the first defeat, captured the ark and returned with it to their homeland. The Bible continues relating the events by stipulating that each town to which the ark was taken within the Philistine homeland was afflicted by plagues of mice and hemorrhoids. Totally paralyzed by their populations' conditions, the Philistines, like the pharaoh before them in Egypt prior to the exodus, consulted their priests and diviners and concluded, after seven torturous months, that it would be best to return the ark to the Israelites. After that was done, and with much rejoicing by the Israelites shortly thereafter, the ark was neglected. After residing for 20 years in Kirjath-jearim, it was taken by Israel's new king, Saul, to lead his army, once again, against the Philistines. Despite the fanfare prior to the battle surrounding the ark, Saul never bothered to do what most earlier rulers of Israel did prior to going into battle against an enemy, and that was to "consult" it. By both the ruler and the people, the ark remained neglected for Saul's entire reign.

King David removed it from Jerusalem when he fled the city as a consequence of a conspiracy by Absalom. Then, he changed his mind and had his priest Zadok return it back to Jerusalem. But, it was King David who had earlier resolved to construct in Zion a large tabernacle, within which the ark would be placed. In the interim, he agreed to have the Levites oversee the ark's safety.

Festivals to which numerous animal sacrifices were ritually performed, along with the regular feeding and blessing by the priests of the large gathering masses of poor at the slowly reconstructed tabernacle entrance, became a consistent practice during King David's rule. It was with the arrival of King Solomon, the son of King David, that the newly constructed temple was completed. In its interior, a Holy of Holies area was provided for the ark. The ark would contain only the two stone tablets that had the Ten Commandments engraved on them. The Israelites saw Solomon worshiping at the ark more often than he had earlier done, after God had promised him in a dream that he would bestow wisdom upon Solomon. With the destruction of the temple, after Solomon's death, the ark disappeared. Rumors circulated wildly among the Israelites as to its ultimate resting place. One view stated that it had been taken to Babylonia. Another view claimed that it had never been taken out of Jerusalem; instead, it was hidden beneath the temple, in a dark area where wood for the temple was stored: all theories that appeared to be driven by idle priests, prone to gossip and conjectures. Whether it was a ruse consciously created by some of those priests to distract the public's knowledge as to its actual destination, in time, possibly will be determined by continued and determined archaeological and historical research and further acquisition of material and written evidence.

The important matter is that it's at this precise juncture in history that Ethiopian Christian Copts strongly held, and continue to hold, to the view that King Solomon and the queen of Sheba's son Prince Menelik I was coopted by one of Solomon's high priests. He was informed that the ark was to be removed from Jerusalem and needed a safe sanctuary far away—the "ideal" location, Ethiopia. Another rendition of Menelik I's role in the disappearance of the ark was that he knew nothing of its acquisition by one of Solomon's old priests and was only informed about it a few weeks after his departure from Jerusalem, on his way back home to Ethiopia. In either case, he acquired the ark, and his duty as Solomon's son was to find a safe sanctuary for it, immediately upon his arrival. Another version as to the whereabouts of the ark has been claimed by a few Muslim historians; the prophet Mohammed's cousin 'Ibn Abbas, the earliest founder of Quranic thought and clarification, has been attributed with the view that the ark with Moses' Rod (instead of it being referred to as belonging to Aaron) lies in the Lake of Tiberius, and that it will be restored on the last day of creation. How the ark initially found its way into the lake, historically, seems quite vague. But it seems that numerous theories abound, in Judaism, Christianity, and Islam. Where the final resting place of the ark actually is, from the archaeological and historical evidence currently available, will require much more excavation and precise documentation. But, in the interim, it appears that the pitched battle that had existed between the Beta Israel lost Jewish tribe in the northwestern highlands of Ethiopia and the majority population of Ethiopian Christian Copts colored and played a significant role in the insistence by

Worshipers gather for Lent at the Church of Our Lady Mary of Zion in Axum, Ethiopia, where some say the Ark of the Covenant is housed. (Franco Taddio/Stockphotopro)

Ethiopia's Christians that the ark actually found its way over the past 2.5 millennia to Axum.

With the legendary folklore that circulated *within* the majority Ethiopian Christian Coptic population from the 14th to the early 17th centuries regarding the courage and ferocious fighting that members of Beta Israel were able to withstand from all Ethiopian Christian advances on them, the church elites began to fear the overall ideological impact of Beta Israel on their extensive Christian following. Numerous ways by which they felt they could attempt to neutralize their growing influence within the ranks of their Christian followers was to tone down a few notches the aggressive rhetoric against Beta Israel, which had evolved unabated over the past two millennia. They allowed Beta Israel to acquire more work as artisans and tenant farmers. Beta Israel became identified, in the minds of the majority Ethiopian Christian population, as being very skillful carpenters and masons, who played an extremely important role in the construction of local and regional churches and palaces. Most of these changes occurred during the thriving Gondar dynasty. This, in time, allowed many members of Beta Israel between 1632 to 1755 CE to settle the urban areas. Beta Israel grew much more influential during that period, in the large Ethiopian Amarinia and Tigre-speaking communities. Increasing land grants were awarded to Beta Israel. A parallel track, which was also pursued by the elites within the church, was to not only attempt to neutralize Beta Israel's growing influence in the minds and hearts of

the church's Ethiopian Christian faithful, but, most significantly, to appropriate the most holy of holy symbols—the ark. The logic being that, with the acquisition of the Ark of the Covenant as the ultimate prize that any Jewish or Christian group could dream of "owning," then, according to the elites within the church's thinking, it would not only stop the "bleeding" of many of their Christian faithful to the alien intrusive presence of the Beta Israel but, most importantly, it would *legitimize* their church *forever* in relation to Beta Israel, or for that matter, any other non-Christian or up-and-coming "renegade" religious group within the church. So, therefore, in their minds, that would justify any false and inaccurate claims that would be made as to the *genuine* ownership of the Ark of the Covenant, or for that matter, toward what they saw as their potential real rivals for such a claim, the Beta Israel. To permanently silence the Beta Israel, by 1755 to the late 19th century, Ethiopia's Christian Copts in the Gondar province scapegoated the Jewish tribe as their collective economic and social fortunes declined. That scape-goating technique was the usual boiler-plate sort, which has often been used with precision and sadistically against vulnerable minorities by dominant majorities throughout world history—single them out and criticize them (question their loyalty, patriotism, and so forth), demonize them, and then kill them. The Ethiopian Christian Copts in the northwestern highlands were only too happy to unload the Beta Israel in the late 1970s, 1980s, and 1990s onto modern Israel, in the airlifts that carried most of them to their new homeland. That way, no group within Ethiopia could ever again contest Ethiopia's Christian Copts' actual direct connection to ancient Israel and, according to them, their justifiable claim to the *sole* ownership of the original Ark of the Covenant.

With that in mind, therefore, the ark placed in Axum, within the treasury building, could be any representation of the original ark, as could be found in most synagogues during their ritual religious services throughout the world. It would be taken out in a few religious processions so that it would be exhibited to the faithful, and it would indirectly convey to the Jewish and non-Jewish world that God made certain that Ethiopia's Christian Copts were the *actual* descendants of his chosen people through Menelik I, the son of King Solomon and the queen of Sheba. It would then be safely hidden behind the veil in the church and an actual "smoke screen" would be created by a few of the priests, which created a sense of awe among the uninformed or possibly an illiterate segment of the parishioners. They would regard that smoke spewing forth from behind that veil as the actual presence of the Lord within their house of worship—as had been mentioned in the Old Testament. Some earlier researchers who had a quick glimpse of the purported ark behind the veil in Axum concluded that, as a result of their expertise with ancient and medieval artifacts and treasures, the ark in the Church of Our Lady Mary of Zion did not appear to date back any farther than the early to mid-14th century. They felt that the priests and the Ethiopian Christian Copts directly connected with the church were being rather

disingenuous, which would bring the neutral observer to question why there was the need for the caginess and deception. Is it an attempt by this Ethiopian Christian group in its rural northwestern highlands setting to coopt the Jewish faith as a whole (or for that matter, any other faith), and possibly, in their own way, have all Jews and non-Christians convert to their Christian faith—after all, as they would have others believe, who does God "live" with through the ark? Or might there be some other reason(s) behind the need to have others believe that they actually have always had sole ownership of the ark for over the past 2.5 millennia? Clearly, the religious and historical plot continues to thicken.

So, therefore, in the continuing academic and historical searches, debates, and attempts to prove beyond a shadow of doubt the *genuineness* of the ark in the treasury building in Axum, Ethiopia's Christian Coptic church officials should be much more transparent and willing to allow professional scientists and archaeologists in to view the ark and, if possible, be allowed to perform scientific tests that would not in any way disturb the sanctity of the ark but would put to rest permanently any doubts concerning the actual final resting place of the original ark. It would clearly and very strongly not only benefit the scholarly and scientific community but the Ethiopian Christian Coptic church as well. It would forever establish the truthfulness of their claim. The well and vigorously researched scientific data need to be shared with the public, so that the historical narrative regarding the Ark of the Covenant can be verified.

References and Further Reading

Budge, Wallis. *The Queen of Sheba and Her Only Son Menyelek*. London: Martin Hopkinson, 1922.

Burstein, Stanley M. *Ancient African Civilizations: Kush and Axum*. Princeton, NJ: Markus Wiener Publisher, 1997.

Carew, Mairead. *Tara and the Ark of the Covenant: A Search for the Ark of the Covenant by British Israelites on the Hill of Tara, 1899–1900*. London: Royal Irish Academy, 2003.

Curtin, Philip D. *African History*. New York: Longman, 1995.

Curtin, Philip D. *Image of Africa*, Vols. 1 and 2. Madison: University of Wisconsin Press, 2004.

Curtin, Philip D. *On the Fringes of History*. Columbus: Ohio State University Press, 2005.

Hancock, Graham. *The Sign and the Seal: The Quest for the Lost Ark of the Covenant*. Clearwater, FL.: Touchstone Books, 1993.

Huggins, Willis Nathanial, & John G. Jackson. *An Introduction to African Civilizations: With Main Currents in Ethiopian History*. Westport, CT: Greenwood, 1969.

Littman, Sheba Enno. *The Legend of the Queen of Sheba in the Tradition of Axum.* Charleston, SC: BiblioLife, 2009.

Pankhurst, Richard. *The Ethiopian Borderlands: Essays in Regional History from Ancient Times to the End of the 18th Century.* Trenton, NJ: Red Sea Press, 1997.

Robert, Dana L. *Brief History of Missions.* Oxford: Wiley, 2009.

Shillington, Kevin. *History of Africa.* New York: Palgrave Macmillan, 2005.

Tibebu, Teshale. *The Making of Modern Ethiopia, 1896–1974.* Trenton, NJ: Red Sea Press, 1995.

Wilson, Marvin R. *Our Father Abraham: Jewish Roots of the Christian Faith.* Grand Rapids, MI: Eerdman's, 1990.

CON

Between the exodus from Egypt and Solomon's construction of the first temple in Jerusalem, the Ark of the Covenant played an important role in the narrative presented in the Hebrew Old Testament. However, following the construction of the temple in the mid-10th century BCE, the ark's role lessened, and it nearly disappeared from the Old Testament. In modern times, the absence of the Ark of the Covenant has been questioned by some, and an interest has been taken in locating it. Some who have attempted to do so claim that the ark may have disappeared from the Bible's historical narratives because it was removed from the temple and from Judah. The theories on why this is so vary between sources, but two widely supported proposals identify the Cathedral of Maryam Tseyon, or Mary of Zion, in Aksum, Ethiopia, as the final resting place of the ark; however, there is much evidence against this theory.

Story of Theft by Queen of Sheba's Son

The earlier story of the Ark of the Covenant's exodus to Aksum is found in the Ethiopian epic Kebra Nagast, translated as "The Glory of Kings." This narrative concerns the queen of Sheba, found in I Kings 10 and II Chronicles 9, and her visit to King Solomon, ruler over Israel and Judah. According to the Kebra Nagast, before returning to her country, the queen of Sheba conceived a son with King Solomon, after he tricked her into submitting to his will. Although born in Ethiopia, the queen's son, Menelik, visited his father years later and won his favor. Menelik refused his father's offer to stay in Judah and be king, saying he had promised his mother he would return to Aksum. King Solomon, grieved to see his now-favorite son leave, required his advisers to send their first-born sons to Aksum with Menelik to become his aides and advisers. These children of Judah could not bear to leave their country and their religious

An Ethiopian miniature depicting the meeting of Solomon and the Queen of Sheba. (Giraudon/Art Resource, NY)

heritage, so they plotted with Menelik to steal the Ark of the Covenant and bring it to Ethiopia. Their plan succeeded, and along with the ark, they are said to have brought Yahweh (God) and his favor to Ethiopia, thus supplanting Israel and Judah as his home. This narrative shows Menelik as the beginning of an Ethiopian royal dynasty that continues David's family, and through this, the covenant between Yahweh and David.

The legend of the Kebra Nagast has resonated through Ethiopian history for many centuries. Until the end of the reign of Emperor Haile Selassie in 1974, the Ethiopian constitution officially recognized the ruling dynasty as descending from King Solomon and the queen of Sheba through Menelik. Similar beliefs were often noted for other governing figures, said to have been of the lines of those who accompanied Menelik from Judah. Perhaps the most recognized modern effect of the Kebra Nagast is the theory that the Ark of the Covenant remains in the Cathedral of Maryam Tseyon, protected and used annually in the Timkat festival each January. This stance can be found in a tour book written by Girma Elias in 1997, *Aksum*: *A Guide to Historical Sites in and around Aksum*. The Davidic lineage of Ethiopian royalty and the presence of the Ark of the Covenant have both been supported by the Ethiopian Orthodox Church. Through this, the Kebra Nagast has served as a legitimizing national epic behind both the Ethiopian political power and church authority.

Story of Journey to Ethiopia by Way of Egypt

A contrary history of the ark to that contained in the Kebra Nagast was proposed by Graham Hancock (1992) in his book, *The Sign and the Seal: The Quest for the Lost Ark of the Covenant*. Hancock's investigation is framed by his communications with an Ethiopian cleric who claimed to protect the ark in the Cathedral of Maryam Tseyon in Aksum. Various issues of historicity prevented Hancock from embracing the traditional narrative of the Kebra Nagast and led him to develop an alternative route for the movement of the ark between Judah and Aksum. Hancock's theory begins with King Manasseh of Judah removing the ark from the temple. The officiating priests could not bear to destroy the ark or allow it to be lost, so Hancock believes they then brought it to a Judean settlement at Elephantine in Egypt. From there, he proposes, it would have been brought to a group of Jews in Ethiopia, around 470 BCE. The ark remained there until 330 CE, when Ethiopian Christians brought it to Aksum. According to Hancock, this history was remembered but became altered over time and developed into the legend of the Kebra Nagast.

Reasons to Doubt Ark Is in Ethiopian Cathedral

Both of these stories present claims that the trail of the Ark of the Covenant leads to Aksum, Ethiopia. Both narratives present reasons to doubt that the ark now resides in the Cathedral of Maryam Tseyon. Manifold archaeological finds lead to questions concerning the legend of the Kebra Nagast and the presumption that the story told in I Kings 10 and II Chronicles 9 refers to a ruler from ancient Ethiopia. Both the Kebra Nagast and Hancock's theory also contradict the biblical narrative and the modern scholarly interpretation of it.

History of Kebra Nagast Doubted

The most obvious points of contention concern the questions of the historicity of the Kebra Nagast. The primary issue among these is the identification of Ethiopia with the biblical Sheba. Most scholars propose that the biblical Sheba was the historical Saba. The center of this kingdom is recognized as having been in southwest Arabia, in modern-day Yemen. However, some discussion concerning the possibility that Saba, or Sheba, may still have been used in reference to Ethiopia exists. Some authors cite the historian Josephus, who called the queen of Sheba the queen of Egypt and Ethiopia. However, in Josephus's time, the term *Ethiopia* would have been attributed to the kingdom of Meroe in modern Sudan, not the modern nation of Ethiopia. Clearly Josephus's writing cannot be seen as supporting the narrative of the Kebra Nagast.

A further possibility of support comes from a royal inscription that refers to Saba. The inscription appears to be an attempt to lay claim to the rule of the Sabaean people in Ethiopia by a ruler named D'amat. Linguistic evidence and

material remains indicate that the Ethiopian Sabaeans represented in the inscription were closely linked to the larger kingdom in Yemen and may have represented a group of immigrants from the south Arabian kingdom. While these inscriptions and cultural remains show evidence of Sabaeans in Ethiopia, none can be dated earlier than the eighth century BCE. Biblical scholars date the reign of Solomon to the middle of the 10th century BCE. This indicates that the Sabaean people of Ethiopia were likely in Ethiopia too late for the narrative in the Kebra Nagast to be factual. In fact, the earliest writing known today placing Sheba in modern Ethiopia is Michael of Tinnis's *History of the Patriarchs of Alexandria*, written between 1047 and 1077 CE. Thus Ethiopia was likely not known as Sheba during the time of Solomon and probably was not home to a Sabaean culture at the time either. This calls into question the narrative of the Kebra Nagast; if the queen of Sheba was not Ethiopian, the story of Menelik's bringing the ark to Ethiopia does appear to be based on anachronisms rather than historically accurate terms.

One of the key points of importance of the story in the Kebra Nagast is that it shows Ethiopia as a direct recipient of Yahweh's grace and of the religion of Solomon. Tradition states that the ark as well as the Davidic line resided in Ethiopia from the time of Menelik to the present. Thus Ethiopia was supposed to have been a Jewish, then a Christian, nation, under David's descendants since the time of Solomon. However, the Jews of Ethiopia likely did not arrive until the middle of the first millennium BCE. This may have occurred as a result of the neo-Babylonian conquest of Judah, which would have led some living in Judah to flee. This would mean that the earliest Jewish presence would have been in the fifth and sixth centuries BCE. This was a number of centuries after the life of King Solomon in Judah. Clearly this calls into question the legend that the queen of Sheba led her people in converting to Solomon's religion and that her son began a Davidic monarchy in Ethiopia in the 10th century BCE.

Multiple Aksum Gods

Of further concern is the evidence that the rulers of Aksum followed gods other than Yahweh, which contradicts the national legends of Ethiopia. The supreme deity appears to have been Mahrem, recognized in conjunction with Ares. Mahrem was seen as a divine supporter of the Aksumite king and was named as the ruler's invincible father. This would have created a divine monarchy, legitimizing the king with a heavenly patron. Additionally the Aksumites worshiped the local gods Astar and Bahar as well as Ilmuqah, Nuru, Habas, Dhat Himyam, and Dhat Ba'ada from south Arabia. The Yahwist faith that one might expect from the Kebra Nagast was not present in the ruling class of Aksum before the fourth century CE. Clearly the queen of Sheba's or the Ark of the Covenant's presence did not lead to a Judeo-Christian government in Ethiopia in the 10th century BCE.

In about 324 CE, the Ethiopian king, Ezana, began to convert his nation to the Christian faith. Under King Ezana the Ethiopian coins that were minted originally showed a divine symbol composed of a disk and a crescent, just as his predecessors' coins had. This symbol likely represented Mahrem or some other form of religious iconography. During his reign this coinage symbol was replaced with a cross, indicating a conversion to Christianity. Ethiopian Christians also began to travel to Jerusalem on pilgrimages in the fourth century CE. The Ethiopian conversion appears to have been from a faith native to Ethiopia or south Arabia, and not Judaism. It is difficult to support the ark's presence in Ethiopia from the 10th century BCE to the present, as the Kebra Nagast suggests, without an accompanying Judeo-Christian faith arriving through a Solomonic influence.

Portrayal of Queen of Sheba

A further issue concerning the Kebra Nagast is its portrayal of the queen of Sheba. She is depicted as an opulent ruler, able to travel with hundreds of camels, asses, and donkeys, each loaded with gifts for King Solomon. This suggests her power over a kingdom with rich and plentiful resources. Menelik is also presented as a grand ruler. He is accompanied by an assemblage of aides and advisers, who assist in the ruling of Ethiopia. These images of the queen of Sheba and Menelik are intertwined with the Ethiopian legend of the Ark of the Covenant's arrival in Aksum and its continued presence there throughout the centuries. Because of this, it is necessary to ask what the royalty and polities would have been like in ancient Ethiopia, because they may support or call into question the Kebra Nagast's narrative and its ties to the Old Testament.

The earliest evidence for a monarchy in the region comes from a series of inscriptions that names D'amat as ruler over Tigray and Eritrea. One such inscription is the previously mentioned one that also contains the name Saba. The dating of the inscription to the eighth century BCE serves to limit the earliest period of a powerful ruler in Ethiopia. This period saw the emergence of early state-level sites under D'amat, as evident in the growth of settlement sites, which before this were only the size of hamlets and towns. Since this transformation can only be traced to the eighth century BCE, the state-sized sites attributable to the D'amat monarchy could not have been contemporaneous with Solomon's reign and the narrative of the queen of Sheba. Instead, had the queen of Sheba, who lived at the same time as King Solomon, been from Ethiopia, she would have been from a region full of small towns and hamlets. The rich and powerful queen of the Old Testament and the Kebra Nagast very likely could not have been from such small settlements and have offered the gifts these sources attribute to her.

The state-level phase of Ethiopian settlement represented by the D'amat inscriptions came to a close ca. 400 BCE. The following period was marked by the absence of the powerful ruling elite, present during the reign of D'amat. It

---

Haile Selassie and the Rastafari Movement

The Rastafari movement, named for Selassie's precoronation name Ras Tafari Makonnen, holds that Selassie was god incarnate and the 225th monarch in the Solomonic dynasty. He is held as the *messiah* and the reincarnation of Jesus Christ. The movement is a syncretic religion owing much to Judaism and Christianity, but deviating sharply from them as well (though many Rastafari do consider themselves Jews or Christians). Rastafari who identify themselves with the Ethiopian Orthodox Church tend to hold the Kebra Negast in high esteem, and the movement in general particularly reveres the New Testament book of Revelation and the messianic prophecies of the Old Testament.

   Selassie had nothing to do with the origins of the movement, and he neither accepted its claims of his divinity nor condemned the Rastafari. He shared spiritual and political ideologies with the movement, though, and the Rastafari position is that it is not necessary for Selassie to claim to be god, he simply is. Likewise, they do not accept his death, and credit it to a conspiracy or hoax.

---

saw the return to sites the size of hamlets and towns, following what appears to have been a collapse of the D'amat culture in Ethiopia. This style of culture continued until about 150 BCE, which marked the beginning of the Aksumite period. The Aksumite period saw the emergence of a new ruling class and the buildup of sites. The rulers appearing in the Aksumite period, including Ezana, would be those who later kings and emperors of Ethiopia would use as a link to the past and to Menelik and the queen of Sheba. However, there is no archaeological evidence of a connection between the pre-Aksumite cultures and those of the Aksumite period. While the majority of sites show no evidence of a continuation of culture, excavations at Matara have shown a clear break between these two periods. This absence of a connection does more than throw doubt on the official narrative of a chain of rulers. If the Ark of the Covenant had been in Ethiopia during the pre-Aksumite period, the lack of connection between D'amat-period sites and Aksumite-period sites offers no means for the ark to have been protected and passed on from generation to generation across the period of collapse.

   The Kebra Nagast does present attempts to connect the reign of Menelik, which would have been in the 10th century BCE, to the Aksumite period and post-Aksumite period, which began ca. 1000 CE. While Ethiopian tradition held that the nation's ruler was a descendant of Menelik and that his power came in an uninterrupted line from Menelik's first dynasty, it also presented the titles of judges, generals, scribes, priests, and other officials as being those found in the Kebra Nagast for Menelik's entourage of advisers from Judah. However, there is no evidence of these titles being used before the post-Aksumite period. What appears to have happened is that titles used for governmental and cultic officials during the period in which the Kebra Nagast was written were anachronistically

attributed to their equivalent in the Kebra Nagast. This tantalizing evidence of a connection between the period of Solomon's reign and post-Aksumite Ethiopia appears to offer no actual evidence of such ties.

Hancock Theory also Problematic

Hancock's narrative of the Ark of the Covenant's coming to Aksum around 330 CE, after having been held in a Jewish temple at Elephantine, Egypt, until the mid-fifth century BCE, may sidestep various historical issues concerning the story in the Kebra Nagast, but his overall support of the ark's presence in Aksum is based on issues that are logically problematic. The major historical theory Hancock presents to explain the travels of the ark is that he sees little role for the ark in the biblical narrative after the time of Solomon and the construction of the temple. However, there is little reason to believe that the end of the mention of a role for the Ark of the Covenant in the Old Testament story marks its physical disappearance from Judah and Israel.

The initial and probably most important argument against the assumption that the ark vanishes from biblical history and then appears in Ethiopia is that the ark does not disappear from the story at all. II Chronicles 35:4 mentions the ark and its return to the temple. Part of Josiah's reform apparently was to bring the ark back to the temple after it had been removed. Additionally it states that the Levites had been carrying it and protecting it during its absence from the temple. This shows a clear logical flaw in Hancock's argument, because the ark's absence from the Bible cannot be sustained as evidence of its disappearance from Judah. The verse in II Chronicles also fails to support the travel of the ark to Egypt; it presents the ark as having been in the care of the Levites, without any mention of it having left Judah. There is no clear record of the disappearance of the ark from the temple in the biblical narrative, not even by the time of Manasseh, when Hancock posits it was lost. Instead, the Old Testament appears to very easily answer Hancock's concern of what happened to the ark; it was returned to the temple.

It is not possible to wholly argue away Graham Hancock's theory of the ark's lessening importance in biblical history by looking to II Chronicles 35, because he argues that this may represent a conspiracy to hide its loss. However, the Old Testament's theology and history do offer an explanation as to why the ark would have been less important following the construction of the temple or why it subsequently was mentioned fewer times. The ark's original purpose among the new followers of Yahweh was to serve as a representative of their God with them, a throne on which he sat. Before the completion of the temple, the ark was easily brought from one town to another, allowing for the worship of Yahweh to take place over a wide area, without requiring a centralized place of cultic focus. The construction of the temple centralized the worship at Jerusalem.

Thus, after King Solomon's completion of the temple, worship was focused at the temple. The Ark of the Covenant then theoretically played a lesser role, because there was no need for it to travel from one town to another. The temple itself became Yahweh's throne on Earth, from which he protected Jerusalem. As in other ancient Near Eastern cultures, this would also have aided the consolidation of the monarchy, because the control of the temple and the official cult would have become tied with the royal family through the palace.

Issue of Centrality of the Temple

The biblical narrative clearly shows a focus on the Temple and away from other cultic sites, which would lead to a lesser importance for the Ark of the Covenant, but archaeologists have uncovered evidence that these other cultic sites were not neglected and that the temple was only one of many place where Yahweh was worshiped. This may appear to give pause to the explanation that the ark almost disappears from the narrative because of a focus on the temple. While there were many sacred sites at the time of Solomon's construction of the temple and throughout the following centuries, this was not the context in which the narrative was written. The book of I Kings was not written during Solomon's reign, but during the reign of King Josiah, during the sixth century BCE. Josiah attempted to reform the worship of Yahweh in Judah, removing holy sites outside of Jerusalem and focusing all cultic practices in Solomon's temple. When I Kings was written, it was only natural to mirror this new religious system instead of a historically more accurate one. The Ark of the Covenant is missing from the biblical narrative not because it had disappeared, but because the authors sought to reinterpret history in a way that did not require the ark's presence and where the ark may actually have detracted from the theology of the central importance of the temple in Jerusalem.

Hancock's Conspiracy Theory

Graham Hancock's argument for the possibility that the Ark of the Covenant was taken from Jerusalem is based on the presence of gaps in the biblical narrative. He posits that the absence of the ark from history is caused by a cover-up conspiracy to assuage the guilt over the loss of such an important relic. This argument is similar to one used by some who support the Kebra Nagast as an accurate history and who look to the same gap as a reason to believe the ark was stolen. This argument is not only specious, but it also goes against modern understanding of the composition of the Old Testament and the developing theology of the temple. There appears to be no reason to believe that the Bible's narrative does away with the ark, because it is mentioned following Solomon's lifetime and after the reign of Manasseh, the two points where its physical disappearance had been posited.

Argument of Ark's Role in Aksum

Modern supporters of the theory that the Ark of the Covenant is currently in the Cathedral of Maryam Tseyon in Aksum often point to its use in an annual festival and the presence of a clergyman to guard the ark and care for it. Graham Hancock uses conversations with this mysterious guardian as a frame for his novel and as a final proof of its presence. However, there are a number of relatively recent events that throw these modern claims of the ark's guardian into question.

The initial modern point of contention concerning the Ark of the Covenant's survival in Aksum occurs between 1526 and 1542 CE. In this period, Ahmad Gran Muslim Amir of Harar invaded Ethiopia and succeeded in destroying the Cathedral of Maryam Tseyon. Although modern Ethiopians hold that the ark was successfully removed beforehand and hidden, there have been contradictory reports. Between 1769 and 1772, James Bruce, an Irish traveler, visited Ethiopia. He reported in his book that while he was there, the king told him that Ahmad Gran was able to destroy the ark within the cathedral. Bruce later wrote that the king falsely claimed the contrary as well. Because of the contradictory nature of Bruce's account, it is difficult to take either position as the truth, but it is certain that the Cathedral of Maryam Tseyon was destroyed once around the period of Ahmad Gran.

Following Ahmad Gran's destruction of the cathedral in the early 16th century, King Sartsa Dengel led the construction of a small replacement cathedral on top of the ruins following his coronation in 1579. Tradition holds that the ark was returned to Aksum, either for the coronation of King Dengel or soon after his reign. In 1611 the Cathedral of Maryam Tseyon was again destroyed during the Galla War, and again the ark would have been threatened. The modern building was constructed during the reign of Emperor Fasiladas, between 1632 and 1667. Had the ark made its way to Ethiopia at any point before this, it would have had to survive two destructions of the church that tradition marks as its home, as well as other insurgencies and uprisings. This would be quite a feat for a 3,000-year-old wooden box. Taking into account Ethiopia's natural environment, it seems unlikely that it would have been in any condition to survive at all, because the warm moist air of Ethiopia does not preserve wood well.

More Recent Ethiopian Legend

A more recent Ethiopian legend has presented an idea as to how the Ark of the Covenant may have survived these tribulations in spite of the 3,000 years of wear. The theory states that what is recognized as the ark in the Cathedral of Maryam Tseyon is not the Ark of the Covenant, but the inscribed tablet on which Moses presented the Ten Commandments. This stone would be much more likely to have survived the millennia between the reign of Solomon and the present day. However, this myth appears to be a more recent creation, and it

contradicts the Kebra Nagast, which states that Menelik stole the Ark of the Covenant. While this new twist on the older legend may attempt to explain the survival of an item from the 10th century BCE, it suffers the same failings as the theory that the ark was brought to Ethiopia that early. Additionally, if the stories that a stone is now in the cathedral were shown to be true, it would certainly not support the idea that ark had been moved to Ethiopia.

Practice of Ark Replication

A final note belongs to the Ethiopian practice of replicating the Ark of the Covenant. Throughout Ethiopia, many churches are noted as having their own *tabot*, or ark. This term, as noted above, is used in reference to their altar tables, which take on a symbolic meaning similar to the Ark of the Covenant in the Cathedral of Maryam Tseyon. When Graham Hancock watched a ceremony where the ark was taken out of the cathedral by priests and paraded through the streets of Aksum, he said that he knew it was a copy. The guardian of the ark had remained in the Cathedral of Maryam Tseyon, while the *tabot* was carried forth. This was likely a symbolic copy of the ark, similar to those used throughout Ethiopia. There is no reason to make the same assumption as Hancock did that the real ark remained in the cathedral during the procession. It is far more likely that any ark kept there would be a copy. The guardian refuses to show the Ark of the Covenant to anyone, and past guardians have stated that they had only seen it a few times in their lives. Without concrete proof of the ark's existence, such as a viewing and analysis of the materials, it is impossible to present any ark in the Cathedral of Maryam Tseyon as more than a symbol of what the true ark is imagined to be. This position is only strengthened by the historical issues concerning the stories of the ark, the likelihood that the ark would have been destroyed, or at the very least deteriorated by this point, and the changing legend as to the actual nature of the ark. An object is certainly being held today in the Cathedral of Maryam Tseyon in Aksum that Ethiopian tradition has labeled the Ark of the Covenant, but all evidence points to it being a relatively recent creation. And until positive evidence to the contrary is produced, this is the most reasonable conclusion.

References and Further Reading

Elias, Girma. *Aksum: A Guide to Historical Sites in and around Aksum*. Addis Ababa, Ethiopia: Mega Printing Enterprise, 1997.

Grierson, Roderick, and Stuart Munro-Hay. *The Ark of the Covenant*. London: Weidenfeld and Nicolson, 1999.

Hancock, Graham. *The Sign and the Seal: The Quest for the Lost Ark of the Covenant*. New York: Simon and Schuster, 1992.

Hoberman, Barry. "The Ethiopian Legend of the Ark." *Biblical Archaeologist* 46, 2 (1983): 112–14.

Lambdin, Thomas O. *Introduction to Classical Ethiopic (Ge'ez)*, Vol. 24. Harvard Semitic Studies. Winona Lake, IN: Eisenbrauns, 2006.

Mendenhall, George E. *Ancient Israel's Faith and History*, ed. by Gary A. Herion. Louisville, KY: Westminster John Knox Press, 2001.

The Modern Translation of the Kebra Nagast (the Glory of Kings). Trans. by Miguel F. Brooks. Lawrenceville, NJ: Red Sea Press, 1996.

Munro-Hay, S. C. *Aksum: An African Civilisation of Late Antiquity*. Edinburgh: Edinburgh University Press, 1991.

Munro-Hay, Stuart. *The Quest for the Ark of the Covenant: The True History of the Tablets of Moses*. New York: I. B. Tauris, 2005.

Pritchard, James B., et al. *Solomon and Sheba*, ed. by James B. Pritchard. London: Phaidon Press, 1974.

2

The Greek city-states were "democratic" by our modern American definition.

PRO	Cenap Çakmak
CON	John Lee

PRO

Whether the polis, city–state regime of ancient Greece may be viewed as a clear precedent and example of democratic evolution and development is a question that attracts a great deal of attention from analysts and political scientists. Opponents argue that the polis order cannot be taken as a precedent for the current understanding of democracy and its practices because of some of its visible defects and antidemocratic characters. They make particular references to the institution of slavery, which was widely practiced and keenly preserved in the city–states, the polis, implying that it is not proper to speak of a presence of democratic order in a place that relies on such an immorality.

There are other serious objections to accepting the polis regime as the initial form of democratic governance in the modern world. A commonly held objection stresses that this regime may not be seen as democratic simply because it denied equal status to all inhabitants in the cities, recalling that only a small group of people were entitled to cast their votes in the election process. Exclusion of women—and other disadvantaged groups—from the right to vote as well as recognition of wealthy people as true citizens with full entitlements stand out as other major objections. Skeptics, therefore, argue that the methods of legislative action and mode of governance employed in the ancient Greece were not truly democratic.

However, while these objections are valid and address part of the actual picture, a more thorough review and evaluation will reveal that the polis may be in fact seen as a major precursor for the current democratic practice and institutions. To this end, the above objections only refer to some deficiencies of the polis regime and do not constitute sufficient evidence for ruling out its ability to serve as the initial form and first example of democratic rule.

Above all, it should be noted that the opponents fail to consider the definition of the term democracy when staging their objections; they are instead focused on the defective elements of the polis order. But we have to keep in mind that a regime will always be flawed regardless of its achievements and the progress made over time. Who could point to a current democratic regime as the perfect rule

today? Take the U.S. system; despite novel arrangements and visible accomplishments, only one of the two major parties is likely to win the presidential elections. Or consider a number of political systems that introduce election thresholds to attain stability in government and legislation. How is it possible to reconcile the basic premises of democracy in light of the obstacles faced by small parties that are unable to win seats in the parliament because they receive less popular support? Yet we still call the regimes fulfilling the basic requirements a democracy and democratic despite grave flaws and serious defects. Therefore, we must investigate whether the polis fulfills the basic requirements of a democratic rule instead of focusing on its defects.

There are at least three major reasons for regarding the political regime and design as implemented in ancient Greek cities, where direct participation was allowed in the rule of their respective nations, as the forerunner of the modern understanding of democracy. Above all, in order to appraise whether the political regime implemented in the polis of ancient Greece was democratic and the initial form of current democratic regime, we have to rely on an objective and authoritative definition of the term democracy. In addition, we must also identify some basic requirements for such a definition. And if the regime in the polis fits in the framework defined by these requirements and definition, then there should be no objection as to whether that regime was truly democratic.

In this case, the set of requirements identified to determine whether a regime is democratic or not should be employed objectively; in other words, the defects and immoral aspects of the polis democracy should not mislead us to conclude that it is not as democratic, despite meeting the criteria and preconditions for being a democracy. If, in its simplest sense, democracy is government by the people and ruled by the majority, we must look at the political regime in the polis to determine whether these two basic criteria are met.

Second, unfair elements and immoral practices in city–state democracies should not discourage us from seeing this regime as the initial example of modern democracy. Criticizing a regime or mode of governance because of its defects and flaws is different from defining the same regime as undemocratic. Whether a regime is democracy is determined via the basic criteria employed for defining the term, and once it is concluded that a given regime is a democracy, its flaws should not lead us to conclude otherwise.

In fact, modern democracies have suffered from serious flaws with respect to fair representation, greater participation, and equality before the law. But they have remained as democracies despite these flaws and shortcomings. It should also be noted that current democracies—even the most advanced ones— have to deal with similar problems to a certain degree. Despite novel arrangements and progress over time, women's participation in government and legislative bodies is still meager in many countries. To deal with this problem, some advanced democracies rely on policies of affirmative action. Low voter turnout rate—for

instance in the United States—is a big problem that needs to be addressed for fairer representation and a more legitimate rule. These are all real and serious problems; however, we still define these regimes as democratic simply because they meet the basic criteria for being a democracy.

The same should also be true for the city–states of ancient Greece. It is true that not all citizens were entitled to cast a vote in a polis; it is also true that slaves were leading a miserable life, and we acknowledge that views of ordinary people who have a right to vote were easily won over by eloquent speakers. But we still observe the same problems in different forms and degrees. Yet we regard the current regimes as democracy despite these problems. Why should we not do the same with the poleis of ancient Greece?

Third, whether the democratic experience in the ancient Greek cities has influenced emergence of democratic regimes in the following centuries is a matter of controversy that needs extensive scholarly inquiry. Regardless of whether it has served as a source of inspiration, this experience represents the forerunner of the modern understanding of democratic practice. In other words, we still have to rely on the basic requirements employed to define the term democracy when attempting to determine whether the experience in the polis is actually democratic, even if we conclude that subsequent regimes and modes of governance did not follow the path and precedent set by that experience.

History needs to be progressive and linear; and for this reason, developments following the experience in the polis of the ancient Greeks are not necessarily affected by that experience. After all, people might not have liked it and may have wanted to replace it with another form of rule, perhaps an alternative means of rule could have replaced democracy as experienced by the ancient Greeks. Therefore, even if a historical survey concludes that the practices in poleis did not set a precedent for the subsequent generations, this will not necessarily mean they were not governed by some form of democracy.

Defining the Term Democracy

In his famous Gettysburg Address, President Abraham Lincoln defined democracy as "government of the People, by the People and for the People." Since then, this definition has gained wide acceptance by thinkers, pundits, and scholars. Lincoln's seminal definition refers to three key elements of democracy: that democracy derives its legitimacy from the people's commitment to it; that the people extensively participate in governmental affairs and processes; and that democracy actually seeks to realize the common welfare and safeguard the rights and freedoms of individuals.

Obviously the key in this definition is the "people." It is known that the term democracy literally means people's government and that it was created by combination of *demos*, meaning "people" and *kratia*, meaning "government."

The Acropolis of Athens as viewed from the Hill of Philopappus (also known as the Hill of the Muses). The Acropolis contained both the civic and religious buildings of the Athenian city-state. Its position on a high limestone outcrop provided defense from neighboring city-states. (iStockphoto)

It is also evident that the current usage of the notion of democracy has evolved from its Greek original *demokratia*. This implies that the current usage of the notion of democracy may be etymologically linked to the democratic regime as practiced in ancient Greece's city-states.

This practice represents the most direct form of democracy; in this system, all citizens met periodically to elect their rulers and other state officials, enact legislation, and discuss governmental issues. Of course it was far from being perfect; slaves, women, and foreigners were not entitled to cast a vote and express their views in assembly meetings.

Despite the flaws, this direct democracy fulfilled the fundamental criteria referred to above. Especially Athens displayed the great achievements in creating a more fair and just regime that ensured greater popular participation in government and legislation. The Athenian democracy, backed by additional reforms in 460 BCE, heavily depended on the popular assembly as the primary sovereign authority. All governmental decisions were made by this institution or required its approval.

More important, there were no restrictions imposed upon those who wanted to participate in the process of government or legislation; basically, anybody was entitled to debate or propose in assembly meetings. Of course there were

downsides associated with this practice. The assembly was convening frequently; therefore, ordinary citizens were rarely able to attend every meeting; as a result of this, a few leading eloquent speakers who were able to articulate their cause dominated the entire process of legislation and rule-making. Yet, popular participation was so extensive and visible that every citizen entitled to participation in the process attended such meetings and held legislative or administrative positions at least once in their lifetime.

The citizens were also allowed to serve in the military and the judicial system. Every judicial decision could be appealed to a citizens' board. Some officers who held key positions were elected by popular vote, and they—generals and treasurers, for example—could be removed from the office by the assembly. Office terms of the elected actors were brief so that others would have the opportunity to serve in governmental posts. Only a few positions in the military were subject to appointment rather than election.

The idea of direct democracy was perfect in theory; but there were serious flaws in practice. Women were excluded from political rights despite being considered full citizens by law. Men slaves and foreigners were also disallowed to participate in the political process. Most important, despite the novel arrangements, aristocrats were still influential in the government and legislation bodies.

However, it is fair to argue that democracy was at work at least in principle because all of the basic requirements were being met. Political leaders and thinkers, in addition, clearly articulated the general rules for application of democracy. The Athenian democracy paid utmost attention to public devotion and the ability and competence of those who were recognized as citizens.

To this end, it will be useful to recall a definition provided by Pericles, a famous ruler in Athens who consolidated democracy in this polis, for the notion of democracy:

> The administration is in the hands of the many and not of the few. But while the law secures equal justice to all alike in their private disputes, the claim of excellence is also recognized; and when a citizen is any way distinguished he is preferred to the public service, not as a matter of privilege but as the reward of merit. Neither is poverty a bar, but a man may benefit his country whenever be the obscurity of the condition. (Fisher, 1901: 99)

Of course, this does not necessarily mean that the actual practice of democracy in poleis has met these criteria and expectations in full. Nonetheless, popular participation in political processes was extensive and fulfilling; popular choice and the decision by the majority were so crucial in decision making that as famous political scientist Stephen D. Tansey recalls, in ancient Greek city–states, "because the majority of citizens had to be convinced if the community were to act, it seems a very high standard of information and debate was often obtained alongside great commitment and loyalty to the state" (Tansey, 2004: 171).

Direct democracy as exercised in poleis was not ideal or perfect; but it was operational and involved all necessary components for greater popular participation in legislation and decision making. *Ecclesia*, the legislative body that also checks the executive body, was open to males over 18 years old. It was subsequently authorized with a practice of *ostrakismos*, under which citizens may convene once a year to determine who has the tendency to become a tyrant. If at least 6,000 citizens declare someone to be implementing this practice, he would be expelled from the community. This was a mechanism envisaged to protect the democratic character of the polis; undoubtedly, it has been abused, and innocent people were unjustly driven away from their domiciles. But malpractice or abuse does not necessarily mean that the idea or the system as a whole was wrong and undemocratic. We all know that judicial errors are still commonplace despite advanced technology and interrogation techniques.

In addition to *Ecclesia*, Athens also had a larger popular assembly, which solely dealt with law making. It consisted of 500 members; 50 being from each *deme*, smaller geographic districts in the city created for better representation. This arrangement was conceived to ensure fair and equal representation of 10 *demes* in the assembly. The *demes* were actually designed as electoral districts; special attention was paid to make sure that aristocrats would constitute a minority in every *deme* so that democracy would be consolidated further.

Representatives from each *deme* were entitled to chair the assembly for one-tenth of the entire year—or 36 days. Chairing the assembly was made possible via a board formed by these representatives. An Athenian citizen was elected randomly to chair this board every day; therefore, it was quite possible for *any* Athenian citizen to become chairman of the board and the entire assembly.

The Athenian people also elected the chief commander of the city–state, the *polemarkhos*. Ten additional commanders, *strategos*, each being one *deme*, were appointed to serve as assistants to him. All these arrangements were introduced for a better and fairer representation in the political institutions of the city–state. Whether these measures have worked is a different story; what really matters is to determine if these measures and arrangements are strong enough for us to conclude that these create a democratic regime. The answer must be yes.

Pericles introduced further safeguards for extensive and greater popular participation; under *graphe para nomon*, every citizen was vested with the authority to protect the fundamental laws and file a lawsuit in request of annulment of a law on the grounds of unconstitutionality. In consideration of reluctance of poor citizens to seek governmental posts and jobs because of lack of financial resources, Pericles also introduced legislation under which those who attend sessions of legislative bodies were entitled to a certain amount of remuneration and allowance. Every member in the *Ecclesia* had the right to speak and make statements, propose a draft bill, and ask for a secret session. The plenary sessions were inaugurated with a call asking who would like to take the stage for a word.

Pericles' Funeral Oration

If there is one individual who is best remembered for reforming the government of Athens, it is Pericles. He transformed the city–state from an aristocracy to an empire in the brief span of his 40 years in office. First as a soldier and then as a statesman, Pericles built Athens into a place of prominence on the Greek peninsula, turning the city's former allies into subject cities, which paid tribute to Athens for protection from the Persian Empire. Although technically still a democracy, debate has raged through the ages as to how much Pericles was led by the citizenry and how much he led them.

Near the end of his time in office, he led Athens into the Peloponnesian War, which would result in the destruction of the city some 25 years after his death. It was after the first year of the war that Pericles delivered his famous funeral oration, attributed to him by Thucydides, which comments on how the Athenian form of government sets them apart from their rivals.

> Our constitution does not copy the laws of neighboring states; we are rather a pattern to others than imitators ourselves. Its administration favors the many instead of the few; this is why it is called a democracy. If we look to the laws, they afford equal justice to all in their private differences; if to social standing, advancement in public life falls to reputation for capacity, class considerations not being allowed to interfere with merit; nor again does poverty bar the way, if a man is able to serve the state, he is not hindered by the obscurity of his condition. The freedom which we enjoy in our government extends also to our ordinary life. There, far from exercising a jealous surveillance over each other, we do not feel called upon to be angry with our neighbor for doing what he likes, or even to indulge in those injurious looks which cannot fail to be offensive, although they inflict no positive penalty. But all this ease in our private relations does not make us lawless as citizens. Against this fear is our chief safeguard, teaching us to obey the magistrate and the laws, particularly such as regard the protection of the injured, whether they are actually on the statute book, or belong to that code which, although unwritten, yet cannot be broken without acknowledged disgrace. (Thucydides 1914: 121–22)

Source: Thucydides. *History of the Peloponnesian War, Done into English by Richard Crawley.* New York: Dutton, 1914.

In addition to such mechanisms, Pericles also wanted to equip the direct democracy with noble arrangements, including *isonomia* and *isegoria. Isonomia* referred to equal treatment of all before the law; according to Pericles, laws provide the same equality for all in personal affairs. In addition, the laws give equal rights for all citizens regardless of their status or rank to participation in political processes. Only merit shall be considered in appointment to governmental posts or other political positions.

Pericles' *isegoria* seeks to ensure freedom of expression for all citizens. Pericles holds that citizens act based on a unique thinking and reasoning; for this reason, he further believes that the citizens should not only participate in state affairs but also do so in accordance with their views. According to him, democracy is based on pluralism and decisions are taken after lengthy deliberations and discussions over diverse views and approaches.

Do the Flaws of Polis Make It Undemocratic?

Polis democracy is mostly criticized for its flaws; critics make particular references to its failure to attract participation of large groups, including women, slaves, and foreigners, in political processes, with there being domination of a small and privileged group in administration, despite novel arrangements and its impracticality.

Problem of Restricted Political Participation in Poleis

It is argued that polis democracy fails to maintain a democratic rule because it excludes women, slaves, and foreigners from inclusion in political processes. This is a very accurate and legitimate criticism; however, polis democracy may not be declared as undemocratic just because it is a defected system and fails to ensure greater popular participation. It should be recalled that not only polis democracy but also most modern democracies suffer from this problem.

Only males were entitled to participation in political decision making in Athenian democracy; Pericles went even further requiring males be born to citizen parents for such entitlement. The elite members were more prone to marriages with foreigners, while the poor mostly married locals; because of the rule, therefore, a substantial number of elites lost citizenship, thus being excluded from political process. Likewise, women had a lower status in Athens; as a consequence of this status, they were denied participation in political processes.

That said, it should be recalled that a number of modern democracies have experienced similar problems. Thomas Jefferson, one of the leading founders of American democracy, owned some 400 slaves. Blacks were not entitled to cast a vote or participate in political processes in a number of modern democratic countries even as late as 20th century. Likewise, extension of suffrage to women is a fairly recent phenomenon; there were still some European countries where women were not allowed to run in the elections in 1970s.

Despite measures taken to address such problems, a number of significant flaws still remain visible in several modern advanced democracies. Above all, women are not fairly represented in many countries. Interest by women in governmental posts is fairly weak; and even if they develop a keen interest, male domination in administrations and legislations is still prevalent and influential.

With the exception of a few countries that have introduced legislations under which a certain percentage of the legislative posts must be filled by women, most modern democracies fail to ensure fair representation of women in government and parliament.

Another problem with respect to poor popular participation in political processes of modern democracies is the visibly low turnout rate. This is a common problem especially in less politicized societies and some advanced democratic states. For instance, turnout rate is around 40 percent in the United States and 60 percent in many European countries. So this suggests that poor political participation is still a problem in modern democracies, despite measures taken to address this problem. This problem notwithstanding, we never consider calling these regimes undemocratic; the same should be the case with the polis.

It is also true that despite measures taken to ensure participation of citizens from all backgrounds and classes in political decision-making processes, eventually only a small number of people gain access to governmental posts and executive positions. Eloquent speakers as well as wealthy citizens eventually established control over the poor and ordinary citizens. For this reason, direct democracy as practiced in poleis was referred to as some sort of aristocracy.

Because the votes of the citizens were crucial in decision making, convincing holders of the right to vote was also important. Citizens uninformed about the matter under review and discussion had to make up their minds based on what had been said on the stage by the speakers who often relied on a strong rhetoric to get what they wanted. This was often the case because most of the citizens holding the right to vote did not get sufficient information because of time constraints and infrequent attendance in the meetings. This made them rely on the arguments by eloquent speakers. It is upheld that citizens were mostly deceived or misinformed by a small number of eloquent speakers who were specialized in convincing ambivalent crowds. Critics, therefore, argue that a regime where the influence of a small number of people is visible and extensive cannot be regarded as a democracy, suggesting that the polis democracy was not a real democracy.

These criticisms are certainly relevant; but the same flaws are frequently observed in many advanced democracies as well. With a few exceptions, politics is something that only wealthy people would participate in within developed or developing countries considered to be democratic. Even though every citizen is entitled to make political decisions and express interest in governmental posts in democratic countries, this is not always possible because of de facto obstructions and barriers.

The case in the United States is especially illustrative; election campaigns often require large sums of money in this country; for this reason, presidential candidates are focused on attracting more funds to finance their campaign. It is also commonplace to observe that presidential candidates are often well-educated and often rich individuals; this implies that although common and ordinary citizens

may run for presidency, theoretically the presidential post is actually reserved for some privileged political actors. It should be recalled that despite its long democratic experience, the United States first elected a black president as late as 2008.

Or consider the impact of media in the elections held in modern democracies; media effect is strongly criticized because of its determinative influence over the election results. People are misinformed or misled by propaganda or media publications; voters may cast their votes based on false or inaccurate information in modern democracies. This implies that modern democracies are actually no different from the polis democracy, in that governmental and administrative positions are sometimes occupied by advantaged groups, including the wealthy, the well educated, or elites.

It is certainly true that there were inherent problems with the direct democracy in the polis that lead in some degree to impracticality. For one thing, the polis democracy was missing a strong and working institutional setting. The composition of the existing institutions was often volatile, making the decision-making process unstable and fragile. A substantial number of citizens virtually did not have time to attend every session; and even when they did, they were uninformed about what was being discussed and reviewed.

Efficiency of the deliberations held in large areas, *agora*, was also controversial. It was pretty likely for the attendants to get distracted because of the large audience and untidy setting. In such an environment, rhetorical approaches and eloquent speeches were pretty influential in shaping the opinions and views of the delegates present at the meetings.

However, it should be recalled that this sort of direct democracy is still being practiced in the modern world. Even though there is no widespread application of it, some small towns in the United States and districts in Switzerland are ruled by such a system where *all* residents who are entitled to participate in the decision-making process convene to discuss their problems and take the proper measures accordingly. Besides, to make a decision on whether direct democracy is really democratic is not relevant to whether it is practical. In other words, a system does not need to be practical in order to be defined as democratic. Thus, we have to admit that polis democracy is actually the initial form of modern democracy if it meets the basic requirements to be considered so, even if it involves some impractical arrangements.

Did Direct Democracy in Greek Poleis Inspire and Influence Other Communities in Coming Ages?

Whether the regime in ancient Greek cities has influenced other societies and nations in the coming ages matters for the sake of locating a deterministic linkage between the democratic character of this order and the modern understanding of democracy. In other words, if modern democracy is an outcome and

culmination of progress made throughout a process initially started by the Greek democracy, then this would mean that Greek democracy is a true inspirer of modern ages.

We are not so sure as to whether the mode of governance in Greek poleis did actually have such an impact over the nations in subsequent ages. However, the practice of direct democracy—just as it was exercised in Greek city–states—in at least some parts of the modern world must be somewhat of an inspiration. This should suggest that people actually like this form of participation because of its directness and ability to ensure fairer representation.

But the crucial question is of course whether representative democracy, the modern practice prevalent all around the world, is an evolved and tailored version of direct democracy in Greek city–states. This is a challenging question that requires a great deal of scholarly effort and extensive research.

However, even if we assume that there is no link whatsoever between the modern development of democratic order and the primitive form of government in Greek city–states that lacked a strong institutional setting, this would not necessarily mean Greek poleis were not truly democratic.

Above all, it should be recalled that democracy was only limitedly practiced in ancient Greece. Not all city–states relied on direct democracy as a form of government; Athens appeared to be a leading example of direct democracy. It is generally held that some of its rulers—Solon and Pericles—played the greatest role in the consolidation of direct democracy in this city–state. Their reforms helped a democratic form of governance to emerge; however, they met with serious opposition and reaction from circles with aristocratic tendencies and ambitions.

It is also worth recalling that democracy was only briefly experienced and practiced in ancient Greece. There are some obvious reasons for the collapse of the democratic order in the poleis. The primary reason appears to be the opposition by some philosophers and thinkers because they upheld that it was simply too dangerous and illogical to leave the task of government to the hands of ordinary people.

What is more, defeats in the wars led to a conclusion that suggests democracy was to blame and that if the city had been ruled under an aristocratic regime, they would not have had to deal with the dire consequences of these defeats. All these factors eliminated democratic regimes in ancient Greece, leaving no model for near future generations.

References and Further Reading

Fisher, George Park. *Outlines of Universal History*. New York: American Book Company, 1904.

Hansen, M. H. *The Athenian Democracy in the Age of Demosthenes*. New York: Oxford University Press, 1987.

Hignett, Charles. *A History of the Athenian Constitution*. New York: Oxford University Press, 1962.

Ober, Josiah. *The Athenian Revolution: Essays on Ancient Greek Democracy and Political Theory*. Princeton, NJ: Princeton University Press, 1998.

Ober, Josiah. *Mass and Elite in Democratic Athens: Rhetoric, Ideology and the Power of the People*. Princeton, NJ: Princeton University Press, 1989.

Ober, Josiah, and C. Hendrick, eds. *Demokratia: A Conversation on Democracies, Ancient and Modern*. Princeton, NJ: Princeton University Press, 1989.

Robinson, Eric W., ed. *Ancient Greek Democracy: Readings and Sources*. Hoboken, NJ: Wiley-Blackwell, 2004.

Sinclair, R. K. *Democracy and Participation in Athens*. New York: Cambridge University Press, 1988.

Tansey, Stephen D. *Politics: The Basics*, 3rd edition. New York: Routledge, 2004.

CON

The history of the Greeks was one of ascendancy to heights of intellectual and scientific achievement that was rivaled by no other ancient world civilization. During the Golden Age of Athens, the Greeks were well known for their philosophical, scientific, and cultural achievements. The source of this achievement may never be known completely. Some historians believe that the source of Greek achievement was due to their subjugation of slaves as workers for their glorious civilizations, the enslavement of women, and the promotion of Greek militarism as well as Greek elitism as the forces for their glorious cultural achievements. The purpose of this section is to show that the Greek achievement was founded on the use of slaves, the enslavement of women, or the promotion of militarism and elitism in the Greek society.

Slavery's Role in the Greek Society

As a factor in the rise of Greek culture, slavery has been presented as crucial to the development of leisure and comfortable living that allowed Greek men to cultivate the intellectual life of the Golden Age of Greece. The slaves worked hard to mine the silver in the earth to produce the wealth that allowed for the comfortable conditions for intellectual pursuits of the scholars of the Greek revolution. The historical origin of slavery cannot be ascertained readily, because of the lack of written materials. Furthermore, the existent materials such as written plays and Greek literature only describe them in commonly accepted stereotypes and do not give detailed description of the true state of slaves in Greek society. Historians believe that there is a mention of slavery

in Homeric writings in 1200 BCE. At this time, there is a description of slavery as a form of private property ownership. So, a chieftain could own about 50 slaves per household in a typical Greek family. In the same Homeric writings, there are stories of slave-raiding parties, with their war enemies taken as treasure or reward of their war exertions. But the Homeric writings make no mention of a slave trade or slave dealers. It seems that we may have concluded that slave trafficking was not even existent at the time. In the Homeric time, slavery seems to be a limited institution and not very

Greek capital with slaves, Corinth, Greece. (Scala/Art Resource, NY)

widespread. We also observe that masters were kind and treated slaves well and slaves were trustworthy and loyal to the master (Lloyd 1988).

Later, slavery became an institution in Greece as well as a way of life. Slaves usually originated from various sources, such as one being in debt to someone and working the debt off for a period of time to the master; as those the various Greek societies captured during their war campaigns against one another's cities; or those who were born into this status as children of slaves. Furthermore, some slaves were orphaned or left to die as babies, but then were rescued and later turned into slaves for the people who had raised them. In addition, when a family needed money, they would sell their daughters into slavery in order to survive off the price for their children; sometimes children were kidnapped and placed into slavery for the support of Greek living.

The Greek slaves played an important role in Greek society and their status was dependent on how much labor they contributed to the society. For example, Greek slaves could work in the house being managed by the women of the house. They could participate in the family rituals and sacrifices, but they were limited in their political participations in the government. These household slaves were considered higher-class slaves. However, they could not enter the gymnasiums (schools) or the public assembly where the political discussions took place. Moreover, they were not considered citizens of the city because they were the property of the master. These Greek slaves were also part of the treasury and Athenian police force. The lower classes of slaves were consigned to menial labor, such as a mineworker in the silver mines, where the life expectancy was very short.

It was after the Dark Ages that the trading of slavery begins as a monetary transaction where a slave could be sold for 10 minae ($180 US), if he was

healthy and strong. A weaker or older slave could be sold for 1/2 minae ($9 US). The price of slavery was subject to market pressures in the economy. If there were a war or battle, the price would usually go down, because the supply was plentiful, and the price would go up if the supply were low. Thus, the slaves were treated as an economic property subjected to the flow of the marketplace.

Among the general slave population, female slaves were usually the lowest status slaves, because of the bias against females in general. They would usually take care of the house: shopping, child care, wool working, and cooking. They also served as wet nurses for the newborn children or as cooks. They also served unofficially as confidants of their mistresses or their masters. At times, the limits could be taken advantage off by the masters. Sexual abuse and rape were not uncommon. The masters usually destroyed the babies of these unions. The female slaves were not treated very well in some families.

In the period from 800 to 600 BCE, the role of slavery expanded because of the expansion of the Greek city–states in their exploration of the Mediterranean world. The Greek city–states began to urbanize their cities, utilized coinage, and started to focus on handicrafts as a manufacturing process. In ca. 600 BCE, there was a strong impetus to use slaves in these handicraft industries and take them away from agricultural use.

The city–states started to employ them legally in jobs as checking for counterfeit coinage and as temple slaves. They were public slaves, who had more independence than privately-owned slaves. Thus, the slaves were used to help in manufacturing handicrafts and industrial production in the Greek cities, which helped the Greek economy to prosper and supported the city–states.

Some historians believe that slavery was not that important to the overall survival of the Greek society. There were some societies such as the Spartans that allowed a modicum of freedom from slavery as well as rights for the slaves. Some historians have claimed that the Greeks had so much free time because they enslaved many people. These historians stated that slaves were important for family life, business life, and political life of the city–state. Slavery allowed the Greeks to become urbanized and free to pursue other occupations not attached to the land. They allowed the urbanization of the city–state and freed the Greeks to pursue more intellectual pursuits. These historians have claimed that the ancient Greeks might have been unable to pursue their individual interests and achievements without these slaves. As a result, the city–states depended on the enslavement of many people to support their manufacture and handicraft industries and allowed them to survive in the ancient world. In addition, those who had slaves had more free time to participate in direct democracy, because they could take days off from work to take part in the election process. Furthermore, they could participate in the 40 or so assembly elections during the year. Thus, the slaves indirectly allowed the Greeks to participate in their government and produce democracy as well as to govern the city–state.

The Subjugation of Women Supported the City–State

Greek society was male centric both in status and perception of the role men, but the dependency on the subjugation of women was relative. In some Greek city–states, the women were considered lower class citizens. Men had the right to vote, to take legal action, and to own property in Athens. Women were considered vehicles for the procreation of the species and were needed for intercourse for men. The females were raised and nurtured until they could be used for marriage to another male. Women were then required to manage the household and the children that came through their pregnancies. They were assisted by their women slaves in the raising of families and their husbands in the working world. Women usually received their education in their homes, consisting of domestic duties and chores as well as managing their husbands' slaves or economics.

In the Athenian democracy (475 BCE), men had written several stereotypes of Greek women into their acceptable literature. The men were responsible to control and maintain their women's sexual appetite, because in the Greek psychology and philosophy, women were lustful and could not restrain their appetites. Greek men were by nature controlled by reason and were more able to master themselves with this natural ability. Aristophanes noted in *Lysistrata* that men should satisfy and control their women's sexual desire in order to preserve their reputation and also to procreate an heir for the man's property. Aristotle also noted that young women needed to avoid masturbating, because they could not control their temptations as well as young boys could. Such were the prevailing stereotypes of women as inferior and uncontrollable animals during Athenian times.

Even in their living spaces, women were confined and limited in order to prevent illegitimate intercourse between the two sexes, because the males had to make sure their property went to rich and intelligent sons or grandsons. The living quarters of women were separated from the male quarters in the houses. The women were often escorted to places, because women were deemed necessary for men's survival. Many men believed that women could not protect themselves from other men. Furthermore, women were separated from men into separate quarters during their social engagements. The social spaces were very much limited for the women of Athens.

The ideal Athenian women was as such an obedient servant of the husband or father whose major responsibility was to procreate and train children, manage the house, spin wool, weave cloth, and prepare the food. They were as such domestic servants of the husband and showed a continual social role that was on the surface subjugating and demeaning in the modern context. Their role was to allow men the freedom to pursue their democratic and civic responsibility.

The Athenian women had many different positions in the society. Some women were prostitutes, who plied their trade through the streets of Athens. They lived in places that would be the equivalent of today's brothels, and the

law limited the amount that they could charge the client. They were supposed to weave and cook for the brothel owners. Other women were courtesans or party girls that entertained the men with instruments or intelligent conversations. They went to parties or symposiums where they entertained the male guest at these parties. Some of them owned their own homes where they would entertain men. Concubines became mistresses to men and sometimes would be the outlet for men in a social or relationship manner.

Despite the relative dependence on the low status of women, there were some communities, such as those of the Spartans, that allowed their women higher status, as well as some women who attempted to compete with men in the male-dominated society. The Spartans allowed their women to compete and train with the men in their military exercises. They allowed them to wrestle naked as well as compete in the sports competition and the physical exercises. Some of the Spartan women could own property with the men as well as help to manage the finances during their marriage to Spartan men. Women were encouraged to develop their intellect, own more than a third of the land, and they could marry at a later age than their sisters in Athens. Husbands were usually away at military exercises, but this allowed their wives greater authority in the homes.

Other areas of dominance of women were as priestesses in religious ceremonies and cults. In John Breton Connelly's (2007) *Portrait of a Priestess: Women and Ritual in Ancient Greece*, the author gathered several epic, lyric poems, speeches, and epigrams to support the function of the priestess in the religious life of the Greeks. Priestesses were either nominated, purchased the position, or were elected to the position of priestess. They were bedecked with white linen cloth and performed animal sacrifices for the goddesses. Priestesses had legal and financial benefits as well as social respect. They could own property, had freedom from taxation, and were given priority to hear from the Delphic oracle. Their personal safety was ensured, and they could have front row seats at competitions. They could put their seal on documents and sanctuary law. They could charge a fee for their services from the initiate. The priestesses performed 145 religious ceremonies and ruled over 40 cults during the history of the Athens. In this book, women were as much of equal status as the men because of their religious importance and administration of the temple rituals.

In other religious areas, the women had outlets where they could express their lives and emotions freely in religious exercise. The women, who worshiped Dionysus, were known as *Bacchants* or *Maenads*. They would leave their husbands and families to dance with Dionysius and honor the wine god during their festivals. They would even suckle wild beasts, even if they had newborn children, if they were not hunting and ripping apart the beasts. Female slaves could participate in religious rituals such as the Eleusinian mysteries. These were limited outlets for the Greek women to express their freedom and individual energies outside the domestic sphere.

There is evidence that some women could read and write as well as discuss the current issues of the day. Vase paintings suggested that women could gather and discuss various issues. Women, however, did not socialize with the men if they were considered respectable women. Women could participate in various cults that allowed socialization and dominance of women such as the Maenads of the Bacchus cults of Greece. There was some advantage for women to manage their property and estates, but women could not sell or dispense of the property without permission of their husbands (or fathers). They were also allowed to receive gifts from others and to inherit property if there was no male heir, but this was discouraged. Finally, women could receive a dowry from their fathers, which guaranteed the marriage if the husband did not want to lose the dowry.

In conclusion, some women were relatively independent and well regarded by men in the society. They also held equal status in the religious rituals of the 40 major religious cults in Greece and could influence the men in their ideas toward life and philosophy. Despite the few examples of women's independence, the subjugation of women allowed men the freedom to participate in political life, and left the other outlets (religious and philosophical) on which the women could spend their energies. The men were able to have freedom to pursue their intellectual quests as well as make civic contributions in the government.

The Intellectual Environment Created by Militarism

Militarism is the doctrinal view of politicians and the military that society should be dominated by ideas embodied in a military culture and heritage centered on the ideals of war. Militarists believe that discipline is the highest social policy and that the social order should support the military. National policy is focused on preparing for military strategy and maintaining war operations. Usually, social oppression follows the enforcement of military order on civilian society. Militarism justifies the use of force in diplomatic and international relations and believes that the civilian people are dependent on the goals of the military, which militarists believe is more important than social welfare. Therefore, militarism is undemocratic and antidemocratic in its outlook and respect of civilian welfare. The entire society's economy and culture are utilized to support the goals and objectives of the military in the society. The national budgets and economy are centered on the realization of military goals and attacking other opponents. Politically, the military will hold two offices: military officers and civilian leaders. Usually, in the democratic government, there would be limitations on the holding of two offices at the same time. Militarism is founded on the premise that freedom of speech and association will be limited, because the society will support military services goals, concepts, policies, and war.

The Athenian army and the Spartan helots pursued a militarily dominant policy of society and political diplomacy. The Spartans were known for their physical

---

The Plague and the Peloponnesian War

The Peloponnesian war between Athens and Sparta lasted nearly 30 years, and during the first year, Athenian spirits were running high. According to the Athenian historian, Thucydides, the following summer the city was struck by a plague that had been spreading throughout the eastern Mediterranean. The Athenians, crowded together inside the city, were easy prey for the plague, which spread rapidly. As a result, somewhere between a quarter and a third of the Athenian population perished, including the city–state's inspirational leader Pericles. Thucydides also caught the plague, but he recovered, and his account is the main historical description of what happened. According to his *History of the Peloponnesian War*, the plague caused a general despair that led Athenians to turn their backs on the gods, which might be argued to be a contributing factor to the city's destruction by the conclusion of the war.

---

training of their male and female youth in sports and military contests. The entire society was focused on training the soldier and the use of armies as part of the militarism of the Spartan society. The Greek society was focused on maintaining the war department of the city–state. They had to defend themselves from the onslaught of foreign enemies as well as take over. As a result, the society was focused on preparing, maintaining, and supplying the defense of the city–state.

Militarism became the central policy of the state, because it had to defend itself through maintaining high amounts of the army and navy. The Greek society invented catapults, fortifications, phalanxes, and the trireme battleship to aid their conquest of land and territory. The Athenians focused on developing their navy by building trireme ships and training sailors to transport their soldiers to war. They built and maintained one of the strongest navies in the Greek world. This navy was used in several wars, such as the Peloponnesian war with the Hellenes. The Spartans utilized their civilian army as their choice of military tactics in producing the phalanxes. The Spartans used helots or militarily trained slaves to man their armies. These helots were Messinian soldiers who had been captured and made military slaves. They would take the children of the helots and raised them in communal camps where the boys and women were trained in tactics of war and hand-to-hand combat. They would eat together, train together, and be educated together so that they would become a united armed force. These helots became hardened under Spartan-like conditions of challenge and combat. They were like the special forces of the army who were specifically trained for battle and combat. Some historians say that the training was so tough that it made the marines look like weaklings. These Spartans were well known for their fierceness and toughness in battle, and they had a reputation of being a militarily oriented society.

These two military societies came into conflict during the Peloponnesian war (431–404 BCE) where they attacked each other during several years of battles and

military struggle. It was fought between Athens and Sparta and their respective allies from the Greek world from Sicily to Istanbul to Crete. The war originated because the Spartans were afraid of Athenian power over the Peloponnesian coast. It continued until Lysander, the Spartan general, defeated the Athenian fleet in the battle at Aeogospotomi in 405 BCE, where it was starved to death by cutting off its supplies. The power of the Athenians collapsed. The war was a struggle between sea power and land power, with Athens dominating the Aegean coast and Sparta dominating the land of the Peloponnesian in the Greek peninsula. It was a struggle because Sparta could not dominate Attica or the territory around Athens, and because Athenians would withdraw into their forts in Athens while being supplied by sea power. On the other hand, the Athenians could not establish bases on the Peloponnesian coast because of the strength of Spartan land power until the war ended at Dellium in 424 BCE. Between the years of 423 and 421 BCE, the Athenian alliance weakened and rallied behind Sparta, because of the defeat at Mantinea in 428 BCE. Finally, in 405 BCE Lysander was able to defeat Athens in a battle at Aeogospotami and surround Athens so it could not get supplies.

Athens continued to flourish during the age of Pericles because of the military might of Athens. Military magistrates who managed the struggle between the two powers ruled the society. However, they were limited in their power by the civilian assembly, which both checked and balanced their power. In fact, it was probably the civilian assembly that sued for peace after the Spartan defeat of 403 BCE. They wanted to rest from the continual warfare and depletion of their prosperity. The building projects of the Athenians were contributed from the rich property owners or aristocrats. However, the poor also had a say in the decision over which building to construct and who to hire for their construction. However, the peace maintained by the militarism of Athenian battles created the peace necessary for the survival of the city–states.

Greek Elitism and Pursuit of Excellence of Greek Culture

The idea of Greek elitism is based on the belief or philosophy that the views of those members of an elite or select group—with outstanding personal abilities, intellect, wealth, specialized training, or distinctive traits—must dominate society even if their policies may not support the society as a whole. Elitism also means that the power is concentrated in the hands of the elite or a special class of people. Elitism has the special characteristics of long-term training or study in a particular discipline or practice; a long track record of skill in a particular art or profession; a history or background in a discipline such as military or martial arts; and having great wisdom in a particular field or discipline.

The Greek elitism can be found in the emphasis toward excellence and virtue in their philosophy and worldview. This elitism is founded on the striving

after excellence and training in the arts, military, and political culture of the Greeks. They call this concept *arête* or civil excellence as a person or society. Plato's *Republic* and Aristotle's *Ethics* discuss that the ideal attitude toward life is to strive for outstanding qualities in its leaders and civilian population. An oligarchy ruled some of the city–states, where the power rested in a small segment of elite and educated families. Many argued that wealthy Greeks ruled behind the Greek government, controlling policy with their wealth and influence. They were an exclusively powerful segment of society that ruled the rest of the society with their power over the economy of the state.

Even in Spartan government, this view of elitism was supported by the enactment of a constitution that gave representation to an upper class of Spartans, but also eliminated a life of luxury for the society. Lycurgus, the Spartan founder, created a system of government that had two kings, five *ephors* or executives, a council of 30 elders, and a general assembly comprising all male citizens. Full citizenship was given to an elite known as Spartiates who fought wars for the society. The Spartans also limited luxurious imports from foreigners and discouraged the ownership of private property as well as democratic ideas, but this created equality among Spartans because all had life and status in common to support the state. As a result, an elite could not develop to rule Spartan society, because the citizens supported the state completely, and not any elitist social class.

According to some historians, the Athenians lived modestly and did not have many luxuries. The economy was dependent on maritime trade and agriculture, but most of the food was imported from outside. The cultural achievements were supported from money from the Delian league, which was maintained by diplomacy and Athenian naval power.

Athens gave equal status for the poor through their first government created by Solon in 594 BCE. Solon created a new constitution that attempted to mediate social conflict between the poor and rich in the sixth century BCE. The reforms he enacted were intended to relieve financial burdens by cancelling the debts of the poor and destroying the law of mortgages for the rich. He also allowed access to political participation for the poor, which had depended on the amount of property and the birth status of the candidate for office. The lowest class was called *thetes* (laborers), who could take part in the general assembly, but not run for office. Solon also banned the export of agricultural products except for olive oil. He also offered to abolish their system of weights and measures for a universal system adopted by the other countries in Mediterranean. Solon constructed a supreme court manned by archons or magistrates elected from the people in the assembly. These laws allowed the foundation for democracy of the lower class to develop during Pericles's times.

During Periclean times, the government was structured with 10 generals who were elected by the citizens. In Athens, the government was run by 10 *strategoi* or generals who were elected by 10 clans to conduct military exploits,

receive diplomats, and direct political affairs. The magistrates comprised the next level of power; they were elected every year to do administration tasks for the government functions, such as police. The Great Assembly was an assembly of citizens who were elected to cast votes on various laws. There were about 6,000 elected citizens who were able to vote on various issues. They were able to pass a law with legal immunity. Finally, there was a Council of the Boule, which was managed by 500 representatives and ruled on legal procedures and processes. They also ruled over decisions made in the General Assembly and administrative details of the government. This form of government seems like it was conducive to a ruling by oligarchy or the military in the society.

Greek Elites and Pericles's Government

Despite this structure, when Pericles was made *Strategos* in 445 BCE, he initiated several reforms that would make the votes and the rights of the poor citizens heard in the assembly. One of the rules was the allowance of *thetes*, or Athenians without wealth, to occupy public office. He also had a special salary or *misthophoria* that was paid to citizens who attended the assembly so they did not need to be employed elsewhere and could just focus on the political life. With these two reforms, he enabled the assembly to function effectively, gave his people public service rights, and created the first polis or city–state in Greece. However, it was the emphasis on civic virtue and civic participation that allowed the elite of Greece to participate in these democratic assemblies.

The government established the principle of equality in its policy of supporting the poor and giving them equal say in Athenian government, but it was balanced with wise leaders and an elitist class in the government. This could have been exploited because the poor were extremely poor or they had no knowledge of the laws. So the Athenian democracy enacted three policies: (1) give an income to public civil servants; (2) seek and supply work to the poor; (3) give land to property-less villagers; public assistance for invalids, orphans, and indigents; and other social assistance. It was the first ancient welfare systems for the poor that also allowed them to participate in civilian life. Thus, even the poor Athenian civilians could participate in Greek public life regardless of income or wealth, but they were led by wise political leaders in the *strategoi* or military leaders in their participation in government.

In conclusion, the Greek city–states depended for their survival upon their domination of slaves, subjugation of women, the expansive role of the military, and the guidance of an elite leadership. The Greek city–states depended on the labor of slaves to increase the leisure and free time necessary for the pursuit of democratic activities of the Greeks. In addition, the Greek city–states' subjugation of women allowed men the freedom and space to guide and govern the city–states. It also disenfranchised a majority of the population in order to allow

the men to guide the city-states. Furthermore, the Greek city–states needed the military to protect and expand natural resources to support the economy of the city–states. Finally, the Greek city–states were guided by an elite class of educated and wise leaders, despite the spread of the vote to the poorer classes.

References and Further Reading

Canterella, Eva. *Pandora's Daughter*. Baltimore: Johns Hopkins University Press, 1986.

Connelly, John Breton. *Portrait of a Priestess: Women and Ritual in Ancient Greece*. Princeton, NJ: Princeton University Press, 2007.

Finley, M. I. *Slavery in Classical Antiquity: Views and Controversies*. Cambridge: W. Heffer and Sons, 1960.

Lloyd, Janet. *Slavery in Ancient Greece*. Ithaca, NY: Cornell University Press, 1988.

Murnagham, S. *Women and Slaves in Greco-Roman Culture: Differential Equations*. New York: Routledge, 2002.

Rich, John. *War and Society in Greek World*. New York: Routledge, 2007.

Westerman, W. L. *The Slave Systems of Greek and Roman Antiquity*. Philadelphia: American Philosophical Society, 1955.

Wiederman, Thomas. *Greek and Roman Slavery*. New York: Routledge, 2001.

3

The Ogham Celtic script is derived from the Norse Rune script.

PRO	Justin Corfield
CON	Harald Haarmann

PRO

A number of inscriptions survive from early medieval Ireland, and these have a script on them, which have become known as the Ogham Celtic script. The exact reason why they have been ascribed this name ("Ogham") is unknown, but it is thought to have been derived from the Irish word for the mark made by a sharp weapon—possibly as the inscription would have had to have been made with a sharp weapon. However, another possibility exists, that this early script was derived from the Norse Rune script, as there was much interaction between Scandinavia and Britain throughout their histories.

There are a number of spurious theories about the origins of the Ogham script that need to be dispensed with before the positive assertion of Ogham as derived from the Norse script can be made. Unlike traditional inscriptions of the period (and indeed of later periods), the inscription itself is sometimes on the "flat" of the stone, but can also be found around the edges, with a number of "notches" that would have had to have been made with a strong (and sharp) metal instrument. Altogether there are several hundred inscriptions in the Ogham script that have survived—counts by scholars vary from around 400 to 507—most have been completely deciphered. When first faced with the problem, there was some doubt over the decipherment—indeed there were some who queried the nature of the script, but with the help of "bilingual" stones from England, linguists have worked hard on the Ogham script and have been able to identify a number of vowels and consonants and work out the number of differences between them. The result of their work has been a tabulation of an alphabet that consists of 25 characters. It has also been worked out that the writing, on the edges of the stone, has to be initially read upward on the left-hand side of the stone, and then downward on the right-hand side.

The reason for the adoption of the script is in doubt. Originally there was a theory that the Ogham alphabet was designed by the Irish, and it was specifically created by a number of Druids, presumably a small select group, for the sole reason of having letters that would be incomprehensible to people who spoke Latin.

43

At the time of the earlier inscriptions, in the early fifth century, there were many issues over the cultural expansion from Britain—where Latin was widely spoken after 400 years of Roman rule, albeit coming to an end. It was during the same period that Saint Patrick was conducting his missionary work among the Irish—traditionally arriving in Ireland in 432 CE after the mission of Palladius in 431 CE. Even if the dates are not that accurate, it is clear that the Ogham script was being used at the same time as the initial spread of Christianity.

Thus, according to the theory that the Ogham script was a Druidic one, the Irish devised this script with some form of nationalistic intent and came up with what was essentially a cryptic alphabet that would serve as some form of code. This idea was widely voiced by the writers James Carney and Eoin MacNeill, who drew on the early work of Robert Alexander Stewart Macalister (1870–1950), who held a chair at University College, Dublin, from 1909 until 1943, and had been involved in extensive research and publications on ancient Ireland and archaeology in Ireland. Macalister had studied in Germany and would have been familiar with the Runes there. However, he was adamant about the nationalistic origins of the Ogham script, and his famous work on the Ogham script was *Corpus Inscriptionium Insularum Celticarum*, although Robert Welch's *The Oxford Companion to Irish Literature* did note that some of his work was "marred by idiosyncratic theories—most notably the view of ogam [Ogham scripts]" (1996: 323). Carney and MacNeill believed that the script might have links to secret five-fingered hand signals, which were used in Cisalpine Gaul (modern-day northern Italy) from the early sixth century. If this were so, there would be some clear and systematic pattern to the adoption of a script for largely nationalistic (and political) reasons.

Much of the basis for this Druidic invention of the script also comes from the medieval texts *Auraicept na nÉces, In Lebor Ogaim* ("The Book of Ogham") and several others. These clearly state that 25 scholars devised the script, each of whom gave his name to one letter of the script's alphabet. While this is clearly an interesting story and tradition, like many other medieval stories, many feel it has no historical basis. The idea of the 25 "wise men" is similar to many other bardic traditions whereby stories from ancient history have been retold for generations, and although they do become an important part of the local folklore, they are not accurate accounts of historical events. Indeed most linguists see a close association of the Ogham script with trees. This can be seen with the *b* signified by a single stroke, also known as *beithe* (birch), and the *s* with four strokes for *sail* (willow). The link with trees has seen another theory that the script came from marks on tally sticks. Raised by the Swiss-born Rudolf Thurneysen (1857–1940), who made a detailed study of the early Irish language, this has gained a number of adherents.

Although it is possible that the script might be in some form of purposely designed code, especially given the shortage of writing equipment during the period, many scholars have criticized this theory. They feel that the script is essentially a

transliteration system in which the Irish have managed to put their words into a script, and that the sounds made in the Irish language were so unlike those in Latin that a new script was more appropriate. This has led scholars to compare the Ogham script with the Younger Futhark, or Scandinavian runes, with which there are some clear similarities.

Charles Graves (1812–1899) raised the first ideas about the link between the Ogham script and some Germanic and Nordic runes. A mathematician and also the bishop of Limerick, he was born in Dublin, the son of a lawyer who became the chief police magistrate of Dublin. After a brilliant academic career, Graves became a professor of mathematics at Trinity College, Dublin, in 1843. An Anglican, he was made dean of Castle Chapel, Dublin, in 1860, and four years later was appointed as dean of Clonfert. Two years later he was appointed as bishop of Limerick, Ardfert, and Aghadoe, becoming one of the last bishops appointed before the disestablishment of the church of Ireland. However, he held the position for 33 years, until his death. Working on many mathematical problems, Graves became interested in Irish history and the Brehon laws, which form much of the basis of the Sister Fidelma murder mystery stories published in the 1990s and the 2000s. This led him to conduct a study of the Ogham script, and Graves traveled around making notes of inscriptions and checking on others. It was through his work that much of the decipherment was carried out, and Graves himself was able to compare the inscriptions with Nordic runes, noticing a large number of similarities in their style and the method in which they were written. As an eminent scholar on a range of topics, he lectured frequently about them.

Stone bearing an inscription in Ogham, County Waterford, Ireland. (Michael Carter; Cordaiy Photo Library Ltd./Corbis)

To draw any conclusions about the Ogham script, one of the most important areas to investigate is just where the surviving inscriptions are located, what they contain, and what their purpose is believed to be. Certainly it is highly likely that there were many more inscriptions that have not survived, and it is also easily possible that archaeologists might uncover new inscriptions in the future.

Within Ireland, most of the inscriptions come from Kerry or Cork. Many were out in the open until recently with the antiquary John George Augustus Prim (1821–1875) managing to save those in Dunbell by taking them to the

National Museum of Ireland. Many have subsequently been saved in local museums.

Although most identified with Ireland, there are Ogham inscriptions that have been found in Wales (especially Brecknockshire) and also a number that have been located in some parts of England, the Isle of Man, Scotland, and the Shetland Islands. Damian McManus identifies 382 inscriptions, and there are some for which scholars query whether or not they contain the Ogham script. The University College of Cork lists about 400 inscriptions, many of which were copied by the antiquarian Abraham Abell (1783–1851). From a Quaker family from Cork, Abell was very interested in archaeology and helped found the Cuvierian Society, which was a forerunner of the Cork Historical and Archaeological Society. A linguist of note, he became fascinated by the Ogham inscriptions. R. A. S. Macalister in 1945 recorded 507 inscriptions, and three more had been found by 1949.

The Ogham inscriptions in Wales often include some Latin names, and of the eight that have been found in England, five were in Cornwall and two close to the Cornish border, in Devon. There is also an isolated one that was recovered in Silchester, in Hampshire, and it is presumed to have been written by an Irish settler or perhaps was moved at some later stage. There were three Ogham inscriptions in Scotland—subject of a learned paper by the Scottish antiquarian James Carnegie, 6th Earl of Southesk (1827–1905)—and five from the Isle of Man. Except for the inscription at Silchester, all the other inscriptions were found from around the Irish Sea. All the Ogham inscriptions are on stone, and it seems obvious that they might have been used on paper that has not survived; very few written records survive from this period, earning it the name the Dark Ages. Two appear in later manuscripts—one in the *Annals of Inisfallen* (1193) and the other, a fictional inscription in the *Book of Leinster*, a Middle Irish saga in which reference is made to some texts. It therefore seems likely that the script originated in Cork or Kerry and was later adopted by small groups elsewhere—certainly from around the Irish Sea. This spread is compatible with either the theory that the script comes from the Nordic runes or that it was invented by a small clique of Druids anxious to keep out British influences.

In order for this large number of inscriptions, over a long period of time, to have been in a script that had been invented by the Druids to prevent the Romans and the Romano-British from interpreting them would mean that the society of the period for the region around the Irish Sea—where the inscriptions are found—would have had to have been dominated and controlled by a small elite who were able to dictate the language in use. This might have come about from visiting Druids—certainly the tales of early Christian saints and martyrs of the same period had people traveling extraordinary distances to spread their teachings. It is possible, but the question is whether this is likely. If all the Ogham inscriptions were dated from roughly the same period, again, this would be possible. But some of the inscriptions date from long after the fifth

---

Casting the Runes

As the Christian era unfolded in Ireland, the church banned other written alphabets from use; among these were both Norse runes and Ogham. Associated with witchcraft and paganism, both forms of communication were used in similar ways in Irish popular folklore. One way in which common people sought divine knowledge for life events was by casting the runes, much like the throwing of dice, but with specific rules of determining what the meaning of the different characters might be for the person's future. Just as Nordic runes were used for casting (and still are by modern Wiccans), stones with Ogham script inscribed were also used to invoke protection, good health, and financial success.

---

century—the ones from the Isle of Man date from the 11th and 12th centuries. That a small Druid circle could keep so much control to use a script devised as a code and keep it in use for 600 years seems far less likely.

The similarities between some of the letters of the Runic alphabet and the Ogham inscriptions lend to the belief that the two scripts are related. So did they use a sharp implement to carve them on stone? The idea of the Runic alphabet would have met with some problems in Ireland where the literate might have had trouble capturing the sounds of the Irish dialect or language. Thus the script developed from a spoken language and was devised—as with most other written languages—from a need to convey the spoken language in a permanent form. Indeed many of the Ogham inscriptions mark the boundaries or ownership of land. A much more modern example of this was the Manchu language, which was exclusively spoken until the early 17th century when the Manchus started taking control of parts and later the whole of China. As they moved from a nomadic people to ones involved in the complex work of administration, the Manchu language needed to be written down, and the Manchu writing, from what was originally a Tungusic language, started to be written in a script not that dissimilar from Mongolian. This is only natural, and it seems likely that the Ogham inscriptions had to be introduced as a way of helping with the administration of parts of Ireland and also places farther afield. Indeed the inscriptions largely date from a period when the Indo-European inflectional endings started to be dropped and are, instead, being replaced with a more distinctive series of initial mutations and a considerable variation in consonant quality.

Although many people are mentioned on the stones, there is only one who is known from other sources. This is one from Wales, which refers to Vortiporius, who ruled in Dyfed and is mentioned in Geoffrey of Monmouth's famous *Historia Regnum Britanniae* ("History of the Kings of Britain"), and the stone, which was found at Castell Dwyran, in Carmathenshire, in South Wales, is believed to be his gravestone.

The greatest quandary for many historians was the discovery of the "Ogham Stone" at Silchester in 1893. This has provided scholars with much debate over its origins and why it was found in Silchester, although it is possible that it was moved there at a later date. There were even suggestions that it was a fake, but that theory has been dismissed in favor of the idea that somebody familiar with the Ogham script, possibly from Ireland or Cornwall or Wales might have moved to Silchester and was buried there with the stone marking his grave.

As to the inscriptions themselves, most of them consist of names of people, with the vast majority describing relationships such as "A son of B" (A *maqi* B) or "A from the tribe C" (A *maqi mucoi* C). Most of the surviving inscriptions list people, with some tribal affiliations, and only rarely other information. This means that much of the interpretation over the origins of names comes from the names used. Many of these seem to be personal descriptions of people, with names like "Alive like fire," "Born of the Raven," "Yew of Battle," or "Chief in Battle." These are clearly pagan names, which can be clearly differentiated from the Romano-British names in use at the time. However, a few of the later Ogham inscriptions, which have been found on the Isle of Man or the Shetland Islands, contain what are clearly Viking names. Indeed the 11th-century Ogham inscription found in the churchyard of Kirk Michael has some Viking runes within it, showing a possible relationship between the two.

Although the geographic spread of the inscriptions and what is contained on them do suggest a development of a script similar to Nordic runes, the major argument in favor of this comes from the reason why these inscriptions were made in the first place. The nature of the wording on them suggests that they were either used to plot land ownership or possibly as tombstones. Either way, they were meant to be read and understood by other people. There would be no point in having a field or land demarcated by an inscription that was only legible to a small Druidic elite. The whole purpose of modern signs such as "Private Property" or "Trespassers Will Be Prosecuted" is to stop or dissuade unauthorized people from entering land. This means that they must be clear and intelligible to anybody. And this would be the same with the stones that have carvings in the Ogham script.

Scholars have debated the reason why a number of the Ogham inscriptions are on stones on which, or from which, crosses have been carved. In a number of cases it is clear that the carving of the cross predates the inscription, clearly showing a Christian influence. As mentioned before, the emergence of the Ogham script coincides with the missionary career of Saint Patrick, but the crosses on the stones used for the script show that the use of the Ogham script is clearly not incompatible with Christianity and in fact seems to be able to happily coexist with it.

Furthermore, some of the stones have inscriptions on them, not only in the Ogham script but also in Latin. This is clear on the vast majority of the surviving inscriptions from Britain (but not those from Scotland). As with the Rosetta Stone (which had three scripts), these "bilingual" stones were used to decipher

the Ogham script in the first place. The inscriptions in Latin would be particularly useful for the Romano-British as well as for the learned people in Ireland, so why have them also in the Ogham script? The only reason for this is that there must have been people who could read the Ogham script but who could not read Latin. Thus the marker or tombstone—whatever the purpose of the original stone—would be legible to people, whichever script they could read. It would, therefore, not be hard for some educated people in those days to understand both by simple comparison of the inscriptions, and therefore the idea that the Ogham script was a secret code cannot really be sustained.

So, the absence of evidence supporting other theories about the origins of the Ogham script, the similarities between the Ogham script and the script found on Norse rune stones, and the obviously pagan features on many of the extant Ogham examples prove that the script could not have been Irish or Celtic in origin, but rather Norse.

References and Further Reading

Brash, Richard Rolt. *The Ogam Inscribed Monuments of the Gaedhil in the British Islands: With a Dissertation on the Ogam Character*. London: George Bell and Sons, 1879.

Carney, James. "The Invention of the Ogom Cipher," *Ériu* 26 (1975): 53–65.

Clarke, Amanda, Michael Fulford, Michael Rains, and Ruth Shaffrey. "The Victorian Excavations of 1893." Silchester Roman Town—The Insula IX Town Life Project. Reading, UK: University of Reading, 2001. Available at www.silchester.rdg.ac.uk/victorians/vic_ogham.php (accessed May 30, 2010).

Fulford, Michael, Mark Handley, and Amanda Clarke. "An Early Date for Ogham: The Silchester Ogham Stone Rehabilitated." *Medieval Archaeology* 44 (2000): 1–23.

Graham-Campbell, James, and Colleen E. Batey. *Vikings in Scotland: An Archaeological Survey*. Edinburgh: Edinburgh University Press, 1998.

Macalister, Robert Alexander Stewart. *Corpus Inscriptionum Insularum Celticarum*. Dublin: Stationary Office, 1949.

MacNeill, Eoin. "Archaisms in the Ogham Inscriptions," *Proceedings of the Royal Irish Academy* 39 (1929): 33–53.

McManus, Damian. *A Guide to Ogam*. Maynooth, Ireland: An Sagart, 1991.

Montfort, Paul Rhys. *Nordic Runes: Understanding, Casting and Interpreting the Ancient Viking Oracle*. Rochester, NY: Destiny Books, 2003.

Welch, Robert, ed. *The Oxford Companion to Irish Literature*. New York: Oxford University Press, 1996.

CON

As a topic of cultural memory, the originality of the Ogham (Ogam) Celtic script has always been an agenda of national identity among the Irish. The debate is heated on whether or not Ogham is a Celtic innovation in writing technology or merely an adaptation of Norse runes. As will be seen from the evidence provided, this section contends that the Ogham script is, indeed, a Celtic innovation.

Origin and Distinctiveness of Ogham

Ogham is the oldest of the scripts in which the national language of Ireland, Irish, has been written. Irish is a representative of the Gaelic (Goidelic) branch of the Celtic languages, like Manx (formerly spoken on the Isle of Man in the Irish Sea) and Scottish-Gaelic. As an instrument of writing technology, Ogham carries much more symbolic meaning than the Latin alphabet, which is used to write many languages around the world.

This self-contained connection of Ogham and Irish as an identifier of Celtic culture may be compared to how the Armenian alphabet is exclusively associated with Armenian, the Georgian script with Georgian, or the Hangul script with Korean. Like Ogham, these scripts were devised to serve as a means of written communication in locally specific cultural environments and for speech communities whose languages are clearly distinct from other languages that surround them.

Left: runestone recovered from the from Viking site of Hedeby in present-day Denmark. Right: illustration of Ogham inscriptions on a standing stone from Ireland. (Library of Congress/Green, John Richard: *History of the English People*, 1902)

The popular debate about whether the Ogham script is an original Celtic innovation or a derivation from the runic script gains momentum when considering the symbolic value of Ogham as an identifier of Irish culture. The discussion in this section unfolds beyond the partisanship of revived druidism and beyond the sentimental agenda of Celtic cultural activism. However heated the popular debate may be, there is one instance that can provide a conclusive and substantiated answer and thus clarify the origins of Ogham. This is the scientific study of the history of writing. The arguments presented here speak in favor of the originality of the Ogham script as a Celtic innovation in writing technology.

The proper handling of categorizations, conceptualizations, and terminology poses a special challenge to any assessment of the origins of writing systems. In the reflections about writing systems and their origins—whether scientific or popular—much confusion has been caused by misconceptions about the principles of writing technology and about relationships between individual scripts. To avoid the pitfalls of generalizing categorizations and imprecise terminology, some pertinent issues of writing research will be highlighted in the following, particularly focusing on the principle of alphabetic writing relating to Ogham.

What Kind of an Alphabet Is Ogham?

Ogham is a script that operates with signs that render single sounds. This makes it an alphabet that is a special type of writing system. Alphabets follow the principle of a one-to-one correspondence of sign and sound. Each single letter stands for one individual sound. Historical orthography may blur the perspective. In the spelling of American English, the writing of the word "thru" is closer to the basic alphabetic principle than the rendering of the same word as "through" in British English, where keeping with the spelling an older (medieval) stage in the phonetic development of the English language is still reflected. In most languages using varieties of alphabetic scripts, though, the one-to-one correspondence of sign and sound is much more clearly recognizable than in writing English (e.g., in Spanish, Hindi, Russian, or Finnish).

Alphabets distinguish themselves from scripts such as Babylonian cuneiform or Japanese Hiragana, where a sign corresponds to a syllable (composed of several sounds), and also from logographic writing such as the Chinese characters or the Naxi script in southern China. In logographic scripts, ideas or concepts are rendered with the help of stylized pictures of the things denoted or the ideas described. The simplified Chinese characters of today are highly stylized and no longer disclose their origins as picture symbols. The signs can be visually identified as miniature images in the script used by the Naxi in their ritual books.

The Ogham script belongs to a certain subcategory of alphabetic writing. The Ogham signs render both consonants (such as *p, d*, or *m*) and vowels (such as *a, e*, or *u*) in writing. This sort of writing all the sounds of a language is typical of those alphabets that were or are in use in Europe, for example, the Greek, Cyrillic, Etruscan, Latin, runic, Ogham, and other scripts. These alphabets are different from the varieties of the alphabet that were or are in use in the Near East, the Orient (Saudi Arabia, Iran, Pakistan), and northern Africa, for example, the Phoenician, Hebrew, Arabic, and other scripts. The signs of the latter alphabets only render the consonants—not the vowels—of the languages that are written with such letters.

In the evolutionary history of writing, the alphabet is the youngest type of script that originated in the first half of the second millennium BCE. Before then, only syllabic or logographic scripts had been in use. The "complete" alphabetic writing, with signs for consonants as well as vowels, is the youngest version that was elaborated in the first millennium BCE. The Greeks elaborated the first complete alphabet in the eighth century BCE—as a derivation from the Phoenician script—for writing the Greek language.

There is no prototype alphabet from which all others would have derived. At a certain time in cultural history, writing systems made their appearance in the Near East. These writing systems were organized according to the "one-sign for one-sound" principle. This is true for the Proto-Sinaitic, the Phoenician, Ugaritan, and other variants of local scripts. All these early alphabets are original scripts in the sense that they were crafted—independently—on the basis of the alphabetic principle. Thus, the Phoenicians were inspired by the same writing technique as the people from Ugarit, the ancient harbor on the Syrian coast, without borrowing letter forms from each other.

This process of crafting local scripts by applying the alphabetic principle as a writing technology repeats itself throughout history and has produced a number of unique innovations like Ogham and other scripts (see above). In this context, the notion of "originality" applies to various local scripts, although they represent a common type of writing, and that is sign use according to the alphabetic principle. For instance, the Armenian script that originated in the fifth century CE was inspired by a Syrian version of the alphabet. The Armenian alphabet is categorized as an "original" script, because it introduces a new property, namely a set of indigenous signs that are not derived from any other sign repertory. The same is true for Ogham. This script has been inspired by the alphabetic principle to write Latin. And yet, the Ogham signs are all original and not derived from the forms of Latin letters (unlike runes, which are derived from Latin letters).

Also in the case of some derived alphabets, it is legitimate to speak of an original script, provided it introduces a new property in its system (e.g., the Greek alphabet, which deviates from its source, the Phoenician script, by using both consonant and vowel letters). Conversely, the Latin alphabet is not an original script, because it operates on the same basis (that is, writing consonants and vowels) as its source, the Etruscan alphabet, which, in turn, is an adaptation from the Greek alphabet.

Is Ogham Associated with the Runic Script in Celtic Folklore or Historical Sources?

Even before scholars in the field of writing research clarified the multifaceted trajectories in the history of the alphabet in the past century, there had been valuable knowledge available about old historical relationships between scripts.

This knowledge was preserved in historical literary sources, and it was also reflected in popular tales.

Especially with alphabetic scripts, a certain respect for the origins radiates from the web of local cultural memory. The ancient Greeks knew that the source of their literary tradition was the Phoenician script, and they proudly hailed this cultural heritage. Roman literary sources show that the masters of the Mediterranean were conscious that they had been taught writing by their Etruscan neighbors, a historical truth that has been corroborated by the findings of writing research. The Slavs, who write their languages with the Cyrillic alphabet (i.e., Russians, Belorussians, Ukrainians, Bulgarians, Macedonians, Bosniaks, Serbians), have always known that their script was derived from the Greek alphabet. A similar consciousness about historical cultural relationships has also been typical of cultural memory in India. There is a widespread awareness among educated people from that country that the origins of their alphabetic tradition lie in western Asia.

If Ogham were somehow related to runic, some whatever spurious allusions to that relationship would be expected to be found in the rich Irish literary tradition that goes back as far as the early Middle Ages. In medieval Irish literature, there is no reference to any foreign script from which Ogham might have been derived. Interesting information about the use of Ogham to write on stones and on wooden sticks comes from the early Irish epic "Táin Bó Cúailnge" ("The Cattle Raid of Cooley"). This epic was composed in the seventh century CE; the oldest extant manuscript from which its text is known dates to ca. 1100. In none of the sources do we find any connection between Ogham and runic.

And yet, neither the nonscientific literature nor any popular account about Ogham can be regarded as authentic in regard to the identification of the origins of this script. For this purpose, the findings from writing research have to be inspected with more scrutiny, paying special attention to comparative methods.

Typological Comparisons between Ogham and Runes

There are undeniable similarities in the Ogham and Runic scripts that evoke curiosity as to their relative nature. A historical relationship can be conceived between the two, with the older Runic script providing the source for the younger Ogham. However, scrutiny of the emergence of the two writing systems, their social functions, the composition of their sign inventories, and their techniques reveals that any historical interconnection between the two can be reasonably ruled out. The major similarities are listed below; analysis of these speaks in favor of regional and independent developments.

Separate origins for Ogham and Runic outside Roman state territory

The Runic script and Ogham originated on the periphery of Roman colonial territory in northern Europe. The findings of modern writing research point to the

islands of Denmark (Jutland, in particular) and northern Germany (Schleswig-Holstein) as the area where runic writing was elaborated around 100 CE. Ogham originated on the northwestern fringe of the area of Ireland occupied by the Romans. This island always remained outside direct Roman political control, albeit in contact with the Roman world. The earliest inscriptions in Ogham date to the fourth (possibly to the third) century.

The fact that the two scripts in question were elaborated outside the borders of the Roman state is no coincidence. The prestige pressure that was exerted by Latin—as a language of the state, as a means of intercommunication among all the ethnic groups in the Roman Empire, as a medium of higher education and civilized lifestyles—on local speech communities within Roman-held territory was so strong that any attempt to create an independent script for writing a local language would have been doomed to failure.

This dual aspect of dependence and independence from Latin has been highlighted for Celtic cultural heritage in Ogham: "The richness of Early Irish literature is a paradox: it owes its existence to the example and challenge of Latin, but also to the independence of Ireland from the Roman Empire. If Ireland had been part of the empire, Irish would have had a status similar to that of British [the Celtic spoken in Britain], one local language among many, overshadowed by the immense cultural prestige of Latin" (Charles-Edwards 1995: 736).

The Germanic population that lived within Roman territory became acquainted with the runes but did not use them frequently. While specimens of runic inscriptions are rare in the Roman area, they are more numerous in southern Scandinavia. The Celts of Britain did not use Ogham, and it was also unknown to the Celts who lived in Gaul (modern France). The Gaulish Celts used the Latin alphabet to write their language, although not on a regular basis.

If the runes were the source of Ogham, an exchange of writing technology would be expected through cultural intercommunication between the Germanic and Celtic populations within Roman territory as would the mutual occurrence of runic and Ogham inscriptions in the areas of contact. However, there is no evidence of this. The runes cannot have influenced the emergence of Ogham through later contacts either. When the Scandinavian Vikings reached Ireland (late eighth century CE), Ogham, which had been used there during previous centuries, was in decline.

Separate cultural contacts of Celts and Germanic tribes with the Romans

The knowledge of writing technology, and of the Latin alphabet in particular, in northern Europe was promoted by trade relations the Romans entertained with Germanic peoples and the Celts. Of special interest to the Romans were the trade contacts between Germanic people and Roman merchants in the northeast. There was a prosperous trade of amber from the coast of the Baltic Sea (in the

regions of former Prussia and today's Lithuania), which was in the hands of Germanic middlemen. Trade with the Gaelic Celts from Ireland unfolded across the Irish Sea. These trade relations in the northwest (with the Gaels) and those in the northeast (with the Germans) functioned separately, without any Germanic-Celtic joint ventures that would have made the diffusion of writing know-how from one culture to the other plausible.

As for the contacts between Ireland and Britain, in addition to trade, cultural exchange was vital for "continuing kinship links between the aristocracy either side of the northern part of the Irish Sea during the first half of the first millennium C.E." (Mytum 1992: 29) and must be considered. Social relations between elite families living on both sides of the Irish Sea, which is on both sides of the territorial border of the Roman Empire, certainly furthered idea diffusion either way. The knowledge of alphabet-based Latin literacy was one of the cultural goods Celtic aristocrats on the British side had to offer their kinfolks in Ireland.

Apparently, the local traditions of writing, Runic in the northeast and Ogham in the northwest, were both inspired by literacy in the Roman world, although they evolved under locally specific conditions of cultural contact and independently in the two distinct cultural areas.

The Latin alphabet as a source of inspiration for Ogham derivation for Runes

The inspirational source to base the local scripts among the Celts and the Germanic peoples on the alphabetic principle of writing definitely has to be sought in the Roman world (that is, with Latin), although the ways in which the Latin script inspired the composition of the Runic and Ogham systems differ locally.

The influence of two contemporaneous traditions of writing known in the western part of the Roman world can be distinguished in the runic letter forms: the Latin alphabet and varieties of the so-called Alpine alphabets that had been inspired by Etruscan writing and were used for writing local languages in the Alps (e.g., Raetic, Camunic). Those who elaborated the runic script and composed its sign inventory drew on the graphic material that they found in the Latin letters and in the sign forms of the Alpine scripts.

The composition of the Ogham inventory of signs differs markedly from that for runes in that the graphic shapes of Ogham letters are, in their appearance, completely disconnected from any of the letter forms of the Latin or runic alphabets. In stark contrast to the shapes of the runic letters, Ogham signs are devised as highly abstract motifs and testify to Celtic inventiveness.

So, although writing technology among the Gaels of Ireland was obviously inspired "by the Latin alphabet, the framers of Ogham showed remarkable independence of mind in their choice of script, their alphabetic sequence, and the sounds they chose to represent" (McManus 1996: 342).

The *Book of Ballymote*

Written around 1391, the *Book of Ballymote* is the only source purporting to describe the invention of the Ogham script for the writing of Primitive Irish language, crediting the Celtic god Ogma with the invention of the script. The book was a compendium of different sources on ancient Greek, Roman, Celtic, and Christian knowledge, but one section, the Auraicept na n-Éces, details the invention of Ogham. Up until the Christian era, the script was kept by Druids as a sacred method of communication. Rather than writing the characters on paper, Druidic priests used finger Ogham, or even shin or nose Ogham, to communicate with other scholars. During the Roman period, Julius Caesar, unaware of Ogham's existence, noted that the Druids resisted the art of writing, and instead seemed to memorize large amounts of poetry. It was not until the onset of Christianization that the druidical colleges began to decline and Ogham began to appear in public. The mysteries of the script were gone, however, as it appeared on monuments, usually next to a Latin translation of the text.

The grouping of letters in Ogham (unknown in runic)

The 20 signs of the Ogham script are distributed in four groups with five letters each. The distribution in the stone inscriptions is the following: (1) *N, S, V, L, B*; (2) *Q, C, T, D, H*; (3) *R, Z, NG, G, M*; (4) *I, E, U, O, A*. A fifth group (called *forfeda*, "additional letters"), comprising signs for rendering the diphthongs *EA, OI, UI, IA*, and *AE*, was added later when Ogham was used for writing manuscript texts. The signs to write consonants consist of scores that are incised to the left and right of an imagined stem line or diagonally across it. The stem line is equal to the arris of the stone that is inscribed. Vowels are written by carving notches on the arris).

The organization of Ogham as a system of intentionally designed score-and-notch motifs makes it a true "cipher," that is a secret code. Ogham might have been intended to be a cipher by its very creators, and, as a secret code, it was handed down from one generation of guardians of druid knowledge to the next in secluded circles.

A unique terminology in Irish exists to describe the infrastructure of the sign system and the composition of its graphic constituents:

- *aicme*, "group of letters" (literally "family")
- *feda*, "letters" (plural of the word for "wood").

The Ogham letters have a tally-like shape, suggesting that they might have been transferred to form an inventory of script signs from an older system of reckoning with tally sticks.

- *druim*, "stem line" (literally "ridge")
- *flesc*, "stroke (score, respectively)" (literally "twig").

The convention to group letters, typical of Ogham and untypical of runic, is not unique for the Ogham alphabet but was inspired by Roman grammatical tradition. "It is now generally accepted that the grouping of the letters can be derived from the classification found in Latin grammarians of the first through the fourth centuries C.E." (Russell 1995: 210).

The distinct order of Ogham letters (unknown in runic)

The order of letters in Ogham also differs markedly from that of the runic. In runic, the first to sixth letters form the sequence *f, u, th, a, r, k*. Futhark is the name for the runic alphabet in several local varieties, while the Ogham script is styled the Beithe-luis-nin (after the names of the first, second, and fifth signs).

Separate conventions of naming letters for Ogham signs and runes

Knowledge of the alphabetic principle (via the Latin alphabet as its manifestation) associates the diffusion of another idea, that of the acrophonic principle, which is reflected in both the Ogham and runic scripts. According to the acrophonic principle, the letters are given names using words whose initial sound is the one rendered by the corresponding sign in writing. The acrophonic principle for name giving was known through the works of Roman grammarians, who documented the literary norms for Latin, among them Terentius Varro ("De lingua Latina") and an anonymous treatise "Rhetorica ad C. Herennium"—both dating to the first century BCE—Fabius Quintilianus ("Institutio oratoria") of the first century CE, Aelius Donatus ("Ars maior") of the fourth century.

The names for individual letters in Ogham are associated with such concepts as trees (e.g., alder, birch, oak) or with natural phenomena (e.g., earth, field, rod of metal). Contrasting with this tradition, runic letters are associated with animals (e.g., aurochs, horse), divinities (e.g., Tyr, Ing), and so forth.

The preference for plants for naming script signs (Ogham) in the Celtic tradition is in accord with the famous ornamental style in Celtic art, which is oriented at plants and their parts. Also in the Germanic tradition of ornamental design, there is a certain preference that is distinct from the Celtic. The visual arts among the Germanic tribes are known for their exquisite animal motifs. The animal style in art finds its parallel in giving names of animals to some of the Runic letters.

Different periods of writing on stone for Ogham and runes

Stone was used extensively as a surface for inscriptions for both Ogham and runes. There are significant differences between the Celtic and Germanic tradition, though.

These differences concern the techniques for engraving letters on the stony surface and placing inscriptions and also the time span when stone was used as a writing surface.

Writing on stone with Ogham is a tradition that is separated by cultural chronology from the heritage of rune stones. Stones were inscribed with Ogham centuries before runes were used on such a surface. From the beginnings of Ogham literacy, stone was the preferred surface. The early runic literacy differs markedly, because inscriptions are placed on portable objects. When runes began to be used to write on stone (fifth century CE), the custom of Ogham stone inscriptions had already been flourishing for almost two centuries. Later, when the number of rune stones (inscribed and ornamentated memorial stones) increased in the Germanic cultural domain (i.e., since the seventh century CE), the use of Ogham on stone was already declining. "It has been suggested that the apparent seventh-century demise of orthodox ogam may have been the result of Christian disapproval, rather than replacement by a more effective system [i.e., the Latin alphabet]" (Redknap 1995: 758).

Those who are inclined to look for a historical relationship between Ogham and runes should consider a possible drift of idea diffusion from the Celtic to the Germanic cultural area (and not vice versa) as regards the tradition of inscribing stones. In this context, cultural chronology speaks in favor of the higher age of this medium among the Celts.

Ogham traditional runic texts differ markedly in inscription engraving technique. Ogham signs are written in a way that the edge (or arris) of a stone serves as the stem line on which notches (for vowels) are carved and from which scores (for consonants) are engraved in both directions. In contrast, runic inscriptions appear in straight rows or bands on the surface of the stone, with no regard for the edge.

The Area of Influence and Cultural Chronology of Ogham

Hundreds of inscriptions in Ogham have been preserved on the most durable material, on stone. The texts are found on stones with memorial function, mostly gravestones. The finds concentrate in the southern and central part of Ireland (some 330 inscribed stones) and also in Britain, namely in the southwest and northwest of Wales (some 150 stones). There are only a few dozen stones with Ogham inscriptions from Cornwall, the Isle of Man, and Scotland.

The oldest known engraved stones with Ogham inscriptions are from Ireland, dating to the pre-Christian era, that is to the fourth and possibly to the third century CE. The stones in Britain are younger and date to the early Christian period (fifth century).

All Ogham inscriptions on stone stem from regions with Celtic population, that is, separated from the areas with early settlements of Angles and Saxons.

The chronology of writing technology in the British Isles shows the following stages:

- Latin alphabet (dominating between the first and fourth centuries);
- Ogham (in Ireland in the fourth century);
- Ogham (in Britain in the fifth century), coexisting with the Latin alphabet;
- Ogham (in Ireland from the fifth to seventh centuries), coexisting with the Latin alphabet;
- Anglo-Saxon Runic script (in Britain from ca. 650 to ca. 900), coexisting with the Latin alphabet;
- Latin alphabet (dominating since the Middle Ages, with scholastic Oghams continuing in Irish manuscripts of the post-seventh-century era).

The language of the Oghams on stone is archaic Irish. The texts are monolingual in Ireland, while in Wales, the inscriptions are usually bilingual (Irish and Latin) and digraphic (Ogham and Latin letters). The language of the Ogham texts in manuscripts is Old Irish.

Social Functions of Ogham Contrasted with Those of Early Runes

The original function of using Ogham for commemorative purposes started in the pre-Christian era and continued into the Christian period. The Christianization of the Celts of Ireland dates to the fifth century and is associated with the missionary work of Saint Patrick. Soon after the new religion had spread among the islanders, Irish monks went to Wales and Scotland to do missionary work themselves. They carried Ogham and the Gaelic language in their cultural "baggage." These two markers of Irish identity disappeared from Wales later, and Ogham declined in Scotland as well, but Gaelic continued to be spoken there (i.e., Scottish-Gaelic).

In addition to names, the Ogham inscriptions on gravestones do not contain much information of historical value. The names refer to an individual person and to a family group and are arranged according to a formulaic order: so-and-so, son of so-and-so. These stereotype formulas in monolingual Irish inscriptions find their counterpart in the bilingual texts (including Latin) in Britain.

The case with texts with memorial contents is different:

> The use of Ogham on memorials seems to have given them special power. Some mention deities as tribal ancestors such as Dovinia of the Corcu Duibne. It seems that Ogham stones were important markers indicating property ownership . . . and perhaps these rights were all the stronger through their statement in a magical script. (Mytum 1992: 55*f.*)

In the old days, many Celts believed in the magical properties of the Ogham signs. These beliefs are echoed in the popular Irish tales. Against this background, it becomes understandable why Ogham was also used for writing spells, namely to make the magical cipher infuse the meanings of words and the contents of phrases with supernatural power.

The original functions of the runes differ markedly from those of Ogham. While texts in Ogham appear in a religious context on immobile objects (i.e., stones), runic inscriptions are found on portable objects (e.g., a woman's fibula, a bone comb, spearheads, shafts of various kinds). Most of the early runic inscriptions are very short, containing only one or two words. Of these, one is usually a name, perhaps of the owner of the inscribed object. In the early phase of the use of runes (i.e., until about 400 CE), the range of subjects treated in the inscriptions is quite narrow. Runic literacy of the early centuries is characterized by "the complete absence of any inscriptions dealing with cult, administration, literature, law, and so on" (Williams 2004: 270).

During the Middle Ages (ca. 950–ca. 1150), when literacy in runic flourished among the Anglo-Saxons in Britain and the Scandinavians, Ogham, in its religious function, was no longer in use and lived on only as scholastic Ogham, this being a bookish medium. As such, Ogham was used in manuscripts since the eighth century and in Ireland only. The subjects in the manuscripts are related to the script itself and to its role in Irish literary history. Ogham was taught in schools until the 17th century, and it was never forgotten as a symbol of Irish culture. For written Irish, the only language that was once rendered in the Ogham script, the Latin alphabet has been in use since the Middle Ages.

As this section has shown, there are multiple reasons for arguing that the Ogham script did not derive from Norse runes. Culturally, the influence of the Romans was much more pronounced than that of the Scandinavians. Linguistically, there are more dissimilarities than similarities in characters and organization between the Ogham script and Norse runes. Finally, the script's function within Christian society aligns it more closely with Celtic origins. Bringing all of this evidence to bear proves the point that Ogham was of Celtic derivation rather than Norse.

References and Further Reading

Charles-Edwards, Thomas. "Language and Society among the Insular Celts AD 400–1000." In Miranda J. Green, ed., *The Celtic World* (pp. 703–36). New York: Routledge, 1995.

Elliott, Ralph W. V. "The Runic Script." In Peter T. Daniels and William Bright, eds., *The World's Writing Systems* (pp. 333–39). New York: Oxford University Press, 1996.

Haarmann, Harald. *Foundations of Culture. Knowledge-Construction, Belief Systems, and Worldview in Their Dynamic Interplay*. New York: Peter Lang, 2007.

Haarmann, Harald. *Historia universal de la escritura*. Madrid: Gredos, 2001.

McManus, Damian. *A Guide to Ogam* (Maynooth Monographs 4). Maynooth, Ireland: An Sagart, 1991.

McManus, Damian. "Ogham." In Peter T. Daniels and William Bright, eds., *The World's Writing Systems* (pp. 340–45). New York: Oxford University Press, 1996.

Mytum, Harold. *The Origins of Early Christian Ireland*. New York: Routledge, 1992.

Redknap, Mark. "Early Christianity and Its Monuments." In Miranda J. Green, ed., *The Celtic World* (pp. 737–78). New York: Routledge, 1995.

Russell, Paul. *An Introduction to the Celtic Languages*. New York: Longman, 1995.

Sawyer, Birgit. *The Viking-Age Rune Stones and Commemoration in Early Medieval Scandinavia*. New York: Oxford University Press, 2000.

Williams, Henrik. "Reasons for Runes." In Stephen D. Houston, ed., *The First Writing: Script Invention as History and Process* (pp. 262–73). New York: Cambridge University Press, 2004.

4

The "Trial of Socrates," described by Plato, was an actual event that occurred in 399 BCE, rather than merely a philosophical device used by Sophists in teaching Apologia.

PRO   Todd W. Ewing
CON   John Lee

PRO

In 399 BCE, the people of Athens tried Socrates on charges of impiety and corruption of the city's youth. During the customary daylong trial, both sides presented their case, and in the aftermath the jurors found the philosopher guilty and sentenced him to death. Since the timing of the execution conflicted with a religious festival, in which a ship traveled to the island of Delos to commemorate the return of Theseus, no executions could occur until its return. During the interval, despite attempts to persuade Socrates to escape into exile, he awaited his sentence, which involved drinking hemlock. The ship returned and, amid his friends, he calmly allowed the sentence to be carried out. These elements, on the surface, appear to present a straightforward account of the demise of a prominent citizen of Athens around the close of the fourth century BCE. Yet, in the aftermath of his execution, controversy about his life and teachings prompted accounts from both friends and detractors. From the sources, a complex—and contradictory—image of Socrates emerged, which has prompted scholars to revisit the account of the trial and call into question its historicity. Perhaps, in one sense, Socrates has been placed on trial again.

The critics point to the fact that Plato, one of Socrates's followers, immortalized his teacher in his many philosophical works. Yet, in what has been called the "Socratic problem," in many cases, the line between what Plato advocated and what came out of the mouth of Socrates is blurred. Thus, one is left to puzzle over what, if anything, originated with the master. Further, since Socrates never wrote anything himself, there appears to be no independent source to evaluate Plato's work. The only other surviving contemporary writings are those of Xenophon, whose portrait of Socrates differed from that of Plato; and Aristophanes offered not only another image of the philosopher, but a disparaging one. Aristotle, a student of Plato, commented in a few places on the differences

63

between his teacher and Socrates, but did not provide enough material on the subject to provide evidence for clear-cut answers. Thus, one is left to consider the possibility that Plato fictionalized the account of Socrates's demise in order to create a platform to present various philosophical ideas, such as the nature of justice, piety, and the immortality of the soul. Plato's contemporaries then supposedly seized on this same means to propose and propagate their own ideas, and in the course of time, the myth of Socrates became historical "fact."

The debate about the various elements of his life and thought has multiplied, and the conclusions are varied. Yet, given the impact Socrates has had on subsequent thought and history, one is compelled to wade through the morass and seek some firm answers. Four issues move to the forefront in the discussion: (1) the reality of the trial, (2) the "Socratic problem" and the validity of the sources, (3) an understanding of the actual charge, and (4) the nature of Socrates's execution. No irrefutable answers are available, yet some evidence exists allowing the trial of Socrates to be securely anchored in historicity.

Before addressing the heart of the controversy, one must discuss the reality of an actual trial. Leaving off the discussion about sources and the nature of the charge for now, an examination of the external elements is in order. Meletus, Anytus, and Lycon, men about whom very little is known, prosecuted Socrates for impiety. Plato has his teacher indicate the charge that he is one who "corrupts the youth and does not believe in the gods the state believes in, but in other new spiritual beings" (Plato, *Apology* 24B–C). Xenophon similarly cited the charge as, "Socrates is guilty of rejecting the gods acknowledged by the state and bringing in strange deities: he is also guilty of corrupting the youth" (Xenophon, *Memorabilia* I.i.1). Of note, Diogenes Laertis, a historian of philosophy writing in the third century CE, reported the charge in the same terms (Laertius, *Lives* I.V.40). He asserted that the accusation continued to be posted in the Metroon, a temple that also housed the city's archives. Granted, he received this information secondhand, but given the correspondence with earlier sources, one finds difficulty in believing that he fabricated this claim. Indeed, his citation of Favorinus indicates that he did not simply lift the material from Xenophon or Plato. Furthermore, even opponents of Socrates confirmed the reality of the trial. In 393 BCE, apparently in response to Plato's *Apology*, Polycrates wrote an attack on Socrates, utilizing a fictional speech of one of the plaintiffs, Anytus, to give the "true" reasons for Socrates's guilt. Unfortunately, this work is lost, but many writers in antiquity make reference to the "speech." In his *Apology*, Xenophon, along with other whose works are now lost, countered Polycrates with their own version. Most scholars consider Xenophon's address to the "Accuser" a reference to Polycrates and not the actual plaintiff (Marchant 2002: ix–x). Anton-Hermann Chroust attempted to reconstruct this lost work in 1957, and noted that the responses to Polycrates took issue with his version of the charges and not the actual event (Chroust 1973: 6). Among the other lost works include those by Lysias and Theodectos (Stryker and Slings 2005: 74).

Also, a trial for the cause of impiety was not necessarily a unique event. The friends of Pericles faced charges of impiety at the beginning of the Peloponnesian war. Plutarch records three such impiety trials. Pheidias the sculptor faced charges that he had stolen gold that covered the statue of Athena, but worse, he had included the likenesses of himself as one of the figures on the goddess's shield. Aspasia, Pericles's mistress, also faced the charge because she allegedly supplied freeborn women to him. Anaxagoras faced charges of impiety. Although the sources disagree, the indication is that the philosopher questioned the nature of the sun and moon (Bauman 1990: 37–38). Diagoras the poet faced indictment for impiety in 415 BCE for mocking the Eleusinian mysteries, and Diopeithes the seer is purported to have promoted a law punishing those who did not properly acknowledge the gods or who inquired into the nature of the heavens. Despite some question about these sources, one could agree with Robert Parker's assessment about Diepeithes's proposed law, that "there is no very strong reason to be suspicious" (Parker 2000: 47).

In the end, the writers of antiquity did not question the reality of a trial, nor is the concept of an impiety trial an alien concept. Yet, one could still argue that many of the sources and examples rest on tenuous supports. The sources appear to contradict one another, and one is still left with the "Socratic problem." If the sources cannot be trusted, how can the historicity of the trial be maintained with any certainty?

Among the surviving sources, four are contemporary, or near contemporary, with the life of Socrates. Aristophanes, the comic writer, wrote *Clouds* in 423 BCE, and a later revised version. Among the works of Plato, four works relate to the trial and death of Socrates: *Euthyphro, Apology, Crito*, and *Phaedo*. Xenophon also contributed to the corpus with his own *Apology* and *Memorabilia*. Aristotle was born 15 years after Socrates's death, and although he never knew the philosopher, he spent 20 years at Plato's academy. The *Ethics* provides comments, not about the trial and execution, but about Socrates's teaching, so as a direct source, it can be set aside for now. Later writers and philosophers wrote about Socrates, but when reconstructing the historical man, as mentioned above, these works tend to be placed in the background as suspect. Once again, the problem surrounding the primary sources is that they all seem to present a different picture of Socrates.

Aristophanes presented a character named Socrates, who, unsurprisingly, runs a "think shop." There he rests in a basket suspended above the ground, contemplating the heavens and expounding his views. His students wander about below involved in a variety of studies. The farmer, Strepsiades, in an effort to escape his growing list of creditors, enrolls in an effort to learn the art of making the weak argument strong. He fails in his endeavor and sends his son Pheidippides, who masters the art and helps his father. Yet, in the end, the son turns on him, and he realizes that he has created a monster and thus wreaks vengeance on the school by burning it down. The play associates Socrates with

In this 18th-century engraving, ancient Greek philosopher Socrates holds a cup of hemlock. After being convicted by a court of corrupting the youth of Athens, Socrates chose to be true to his principles by drinking the poison rather than trying to secure a release through bribery or an escape attempt. (Library of Congress)

the natural philosophers and Sophists of the day, who were renowned for their art of rhetoric and charging people large fees for instruction. Aristophanes obviously caricatured his subject, so its use as a means of discovering the historical Socarates can be dismissed; but he presented a portrait of the man that became reality in the minds of Athenians, much like modern sketch comedy parodies politicians, placing distorted images in the public's mind. Plato indicated this truth by having Socrates mention "Aristophanes' comedy" as a distortion of his teaching (*Apology* 19C).

Plato's *Apology*, on the other hand, presents a polished speaker, well versed in the art of speaking. Indeed, despite the traditional claim that he did not prepare a speech beforehand, Socrates presents a formal, "forensic" court speech (Burnyeat 2004). Indeed, the presentation follows all of the proper rhetorical rules. He begins with an exordium to his audience, moves to a prothesis, which states his case, and counters with a refutation. The speech concludes with a typical digression and a final peroration (Allen 1984: 63). This polished urbane Socrates does not match the portrait given by Xenophon, who paints him as a harmless sage dispensing practical bits of wisdom on careers and household management (Wilson 2007: 94). Indeed, even the exchange with Meletus in Plato's version seems to be constructed like a Socratic dialogue (Burnyeat 2002:

Plato's Account of the Death of Socrates

Plato's *Phaedo* contains the death scene of Socrates and presents the scene through the eyes of Phaedo, a close friend of Socrates:

> Crito made a sign to the servant, who was standing by; and he went out, and having been absent for some time, returned with the jailer carrying the cup of poison. Socrates said: "You, my good friend, who are experienced in these matters, shall give me directions how I am to proceed." The man answered: "You have only to walk about until your legs are heavy, and then to lie down, and the poison will act." At the same time he handed the cup to Socrates, who in the easiest and gentlest manner, without the least fear or change of color or feature, looking at the man with all his eyes, Echecrates, as his manner was, took the cup and said: "What do you say about making a libation out of this cup to any god? May I, or not?" The man answered: "We only prepare, Socrates, just so much as we deem enough. I understand," he said: "but I may and must ask the gods to prosper my journey from this to the other world—even so—and so be it according to my prayer." Then raising the cup to his lips, quite readily and cheerfully he drank off the poison. And hitherto most of us had been able to control our sorrow; but now when we saw him drinking, and saw too that he had finished the draught, we could no longer forbear, and in spite of myself my own tears were flowing fast; so that I covered my face and wept, not for him, but at the thought of my own calamity in having to part from such a friend.

Source: Plato, *Phaedo*, Project Gutenberg, www.gutenberg.org/files/1658/1658-h/1658-h.htm. (accessed May 31, 2010)

---

144). Trust in Plato's account is further diminished with the "Socratic problem." Plato, in other works, has obviously placed his own ideas in the mouth of Socrates, showing developed philosophical concepts, such as that of ideal forms. The assumption is, then, that if Plato is guilty in one place of using Socrates as a mouthpiece, he is guilty elsewhere too, and determining the line between the historical and the fictional man is impossible. One must admit that most likely Plato has not presented a verbatim account of the content of the trial speech. To suggest as much shows a complete misunderstanding of classical literary technique. The possibility certainly exists that he has inserted his own ideas into his master's mouth to promote a particular concept. Yet, one is left to wonder about which particular Platonic teachings are set forth in the *Apology*. The purpose of the work does not seem designed to promote a particular philosophical idea, but instead, focuses on vindicating the memory of Socrates. Plato could not have accomplished some sort of didactic goal through a complete fiction in the face of Athenians who actually remembered the trial (Vlastos 1995: 6; Fowler 2005: 65.)

Xenophon's portrait, as mentioned before, is different. Whereas Plato claimed to be an eyewitness, Xenophon admits to having received his information secondhand through the auspices of Hermogenes, traditionally shown as one of the friends of Socrates. The contents of this *Apology* and the *Memorabilia* differ from Plato's accounts. These are presented in more of a narrative form, and he pursues a different agenda. Xenophon sought to create a picture of Socrates as one who faithfully respected the gods of Athens and participated regularly in the city's rituals. One looks in vain for some kind of Socratic harmony, in which the texts could be placed side by side. The *Memorabilia* has been demonstrated to be a response to Polycrates's work and thus, Xenophon has shaped his work accordingly. Some of his assertions about being present with Socrates are suspect and obviously a literary device, but one cannot rule out his claim to have actually heard the teacher on occasion (Marchant 2002: xiii, xiv). Once again, though, one is left with an unsettling uncertainty about the validity of the contents of his work.

Leaving aside the "Socratic problem" for now, what can be known from the sources at hand? Aristophanes, as said, can be dismissed as a concocted image that is not "historical." Plato and Xenophon have obviously fabricated elements and style for their own purposes. Yet these two authors can be mined for a historical "core." Xenophon wrote long after Plato, and he may have known of the latter's efforts to give an account, but one does not find him simply "plagiarizing" (Marchant 2002: x). Various components can be found that correspond in their respective works, showing that each of them was at least acquainted with actual elements that served as part of the backdrop for their own agenda. More significant is that these elements are not essential to their basic argument, and thus there would have been no reason to fabricate them. For example, the charge itself is not something that either would have fabricated. Indeed, both report the charge of impiety and corruption of the city's youth (Plato, *Apology* 24B; Xenophon, *Apology* 11, 12; *Memorabilia* I.i.1).

Other elements correspond. Both make reference to Chaerephon's inquiry of the Delphic Oracle, in which Socrates received his mantle as wisest among men (Plato, *Apology* 21A; Xenophon, *Apology* 14). Both refer to an exchange with Meletus even though Xenophon merely gives an excerpt as opposed to Plato's actual dialogue (Plato, *Apology* 24A–27D; Xenophon, *Apology* 20). Both have him defending against each of the charges, albeit with different emphases. Plato has Socrates refer to the "god" and the "gods" in more vague terms, associating his individual "spirit" with the divine; whereas Xenophon talks about the defendant's association with various city cults. Although the latter likely exaggerates, they both indicate that he addressed this charge (Plato, *Apology* 27D; Xenophon, *Apology* 12–13). The inference that some kind of alternative fine could be paid is reflected in both of these works, even though it is garbled in the tradition, since Xenophon states that he refused to allow his followers to pay (Plato, *Apology*

38C; Xenophon, *Apology* 23). Regardless of the amount or the attitude behind the counteroffer, all accounts recall Socrates making some statement about a fine.

One other element may not only illustrate an association with an actual event in the trial, but serves as an example of how Xenophon extrapolated. In *Apology* 39C–D, Plato "prophesizes" about the "grievous" fate of those who unjustly condemned him. Xenophon related that he prophesized about how Anytus's son, despite advice, would "fall into some disgraceful propensity and will surely go far in the career of vice." This fulfillment is most likely an extrapolation to provide a vindication of Socrates, but reflects an original factor in the trial (*Apology* 29–30). Further, both reference his confidence in facing death. One has difficulty seeing how, or why, both would create an element that does not seem to promote their literary agenda (Plato, *Apology* 41D; Xenophon, *Apology* 27, 32–33). Finally, in connection with his imprisonment, both authors refer to an attempt by his friends to engineer his escape, which he rejected (Plato, *Crito*; Xenophon, *Apology* 23). Once again, they do so without any obvious borrowing from each other, but unconsciously incorporate real life events into their works.

For modern readers, Plato's and Xenophon's literary liberties erode confidence. Yet one must remember that authors in antiquity utilized a different methodology in presenting events and especially speeches. As Thucydides explained in his well-known axiom on the speeches in his work, he simply attempted to capture the essence of what his subjects said, not to provide a verbatim report (Thucydides 2003: i.xxii.1). In the case of the *Apology* of Plato, Charles Kahn noted the distinction between it and the other dialogues in that it is the content of a public speech more in the line of Pericles's funeral orations, thus "quasi-historical" (1996: 88). Doubtless, Plato and Xenophon both placed words in Socrates's mouth. The contrast between the polished Plato and the mercenary Xenophon is obvious, but both seem to have captured the essence of Socrates even as they attempted to exonerate a man pilloried by his own community. They both drew upon their own experience of the man, coupled with recollections of statements he made and attitudes he held. Furthermore, Plato was not being deceitful in his efforts. He certainly based his ideas on some formative teachings of Socrates, to which he felt free to add his own thoughts in order to arrive at what he assumed to be a logical conclusion (Guthrie 1969: 33–34). Yet, once again, this appears to beg the question: If Socrates is a construct of Plato and Xenophon, then, is there any historicity?

One is forced to deal with the "Socratic problem" and whether the historical man can be separated from the literary creation. In the end, no definitive statements can be asserted, but some conclusions can still bring one into the arena of historicity. The usual approach in the search for the historical Socrates is to divide Plato's works into a chronological schema of early, middle, and late. The assumption is that the earlier dialogues would reflect more of Socrates's actual teachings and ideas. Yet, as Plato progressed and matured in his

philosophical thinking, he began to incorporate his own ideas into his Socratic vehicle. Most efforts to separate Plato from Socrates search for stylistic and philosophical developments, which although problematic still offer insights, despite objections (Allen 1984: 8–9; but see Wilson 2007: 101).

Robin Waterfield provides an excellent summary of the four main points used to demonstrate a contrast between Socratic ideas and later Platonic influences. First, there is a contrast between the ideas that one will always choose what is best versus being controlled by one's appetites. Second, Socrates rejects the typical Greek attitude that one could justify harming one's enemies, but in works chronologically identified as later, such as the *Republic*, the idea is actually championed. Third, Socrates limits his inquiries to discovering moral issues, as opposed to a focus on metaphysical explorations of the ideal (which Aristotle seems to associate with his "friends" [i.e., Plato] in *Ethics* I.vi.1). Finally, Socrates seems to proclaim that he himself knew nothing definite as opposed to a later confidence of certainty (Waterfield 1990: 12–13). One must admit that these do not offer definitive proofs about what is distinctly Socratic, but they create a defensible position that a progression of ideas exists in the Platonic corpus. Further, those works that show a mature development of ideas are stylistically linked to works associated with *Theaetetus*, which does offer an external historical anchor. A reference in its introduction is made to the siege of Corinth, which is dated at 369 BCE. Therefore, since the *Apology, Crito*, and *Phaedo* associated with the works are stylistically distinct, the implication is that they are earlier works. One can, with logical consistency, maintain a certain amount of confidence of the existence of a "Socrates" closer to the man himself. With some justification, W. K. C. Guthrie calls the *Apology* the most "Socratic" of Plato's works (1969: 157–58). Although skeptics complain about the inevitable reality that one can never completely separate the Platonic Socrates from the actual man, there is truth in the statement that the *Apology* does not contain any utterance that the actual Socrates could not have reasonably stated publicly (Stryker and Slings 2005: 78).

Plato provided glimpses of the real Socrates, which calls into question the validity of Xenophon's portrait. Along with the *Apology* and the *Memorabilia*, Socrates is the main character in the *Oeconomicus*, a work on managing the estate, and the *Symposium*, which deals with conversations at a dinner party. Socrates appears to be less witty and not as profound in Xenophon (Guthrie 1969: 15). The contrast appears to suggest that one is correct and that the other should be jettisoned in the search for Socrates; and most scholars favor Plato as more authentic. Yet, as Waterfield points out in defense of Xenophon's image of Socrates, a philosopher does not always dwell upon the lofty, but most likely provided occasional advice on practical matters (1990: 19–20). Plus, he did not demonstrate the capability of gleaning the more profound insights of his teacher as Plato could and so dwells upon matters that specifically interested him. Also, where Xenophon does take license, one is reminded once again of the literary

habits of the classical world. Xenophon is utilizing firsthand encounters, anecdotes from his friends, as well as his own limited extrapolations to portray Socrates as a harmless wise man (Wilson 2007: 96).

Guthrie summed up the best approach to the apparent discrepancies in the sources when he indicated that Socrates had to have been a "complex character," with many aspects to his personality. Thus, when each of these writers approached their subject, they did so with their own interests and agenda (Guthrie 1969: 8–9). They did not erase the foundational essence of Socrates and the pivotal events in his life, nor is their literary exaggeration a deceitful fabrication. When one considers other historical persons in antiquity, the same tendencies apply. Jesus of Nazareth, a peasant from Galilee, carried on a teaching ministry until he ran afoul of the Jewish and Roman authorities, who executed him. His followers quickly interpreted his life and work, and not all of them agreed on the details, but the essential historical facts remain. Alexander the Great rose to prominence and conquered the Persian Empire. Those who wrote afterward presented him either as a brute or a visionary. Regardless of how one viewed him, the ancient Near East had become Greek. Harmodius and Aristogeiton assassinated Hipparchus because of a family slight, but later Athenians elevated them as the "tyrant slayers." The motives had been changed, but the essential historicity of the event remains. Likewise, in the case of Socrates, Xenophon and Plato may have colored events with their own style and program, but they did not lose the basic historicity of their master defending himself before the court of Athens, under a charge of impiety.

Another controversy that circulates around the trial of Socrates concerns the motives of the prosecutors. On the surface, Meletus and his associates accused the philosopher of impiety, but this charge seems difficult to define for modern scholars. Also, the suggestion has been put forth that this action served only as a mask for political motivations (Stone 1989: 138–39). Before addressing this view, one must understand the nature of the charge of impiety. Socrates is being accused of not believing in the gods of Athens and of introducing new divinities. The former is not simply a charge of atheism, although this must have been a concern for some of the jurors, but that he did not actually participate in the civic cults, which, being the social creature he seemed to be, is implausible (McPherran 1998: 163–64). Yet, philosophical skepticism is not the only crime of which he was accused. He had not been the first to question the nature of the gods, but the latter part of the charge is more significant. Even in *Euthyphro* Socrates quoted the charge that he is "a maker of gods" (3B) before a reference to the city's gods. He and Meletus were most likely referring to the claim that he had his own divine spirit that gave him advice on occasion (*Apology* 31C–D). Xenophon apparently did not understand the underlying issue and focused on demonstrating where Socrates did participate in the civic cults. Plato, however, seemed to understand the true nature of the charge. Socrates not only

questioned the nature of the gods as worshiped in Athens, but also suggested the existence of a personal spirituality that superseded the state cults. What made him particularly dangerous was that he had a popular following among the children of prominent citizens in Athens.

The modern approach to the trial is to assign political motives to his adversaries. This belief usually stems from Socrates's so-called antidemocratic statements and his relationships with Critias and Alcibiades. Both of these men were associated with oligarchic forces, and Critias stood among the Thirty Tyrants who butchered many in Athens in 404 BCE (just five years before Socrates's own trial). Yet, because of the amnesty in the aftermath of the democratic restoration, charges could not be brought for earlier political crimes. So those seeking to accuse Socrates of political subversion in connection with these off-limit events had to change the charge to a religious one in order to accomplish his destruction. Thomas Brickhouse and Nicholas Smith offer the best refutation of political motivations. In truth, Socrates did make criticisms of democratic practices (see *Apology* 25C for example), but his connection to those who disrupted the city is more to the point. Both Alcibiades and Critias were his known associates, and the assumption persisted that Socrates encouraged them in their destructive activities. Xenophon attempted to address this issue (*Memorbilia* I.ii.12), but once again, he is actually answering Polycrates. Plato's *Socrates*, however, does not mention these associates and the so-called corruption of them; so, if political issues did hang in the background, one wonders why he would not have addressed them a full defense (Brickhouse and Smith 1989: 71, 73–74).

In actuality, one must remember that political motives could have been in the background, but not those suggested above (Parker 2002: 151). Unlike modern Western democracies, Athens did not attempt to separate religion and state. A threat to the state cults equated with a threat to the state. The charge did stem from Socrates's personal divinity, which could undermine the civic cults, and, in the minds of the Athenians, this made Socrates a danger (Wilson 2007: 32–33). In turn, he could continue to teach others to follow his example of undermining the social values of the city (Parker 2002: 153–54). One must remember that the trial took place during a time of social and civic turmoil. Athens had recently lost the Peloponnesian war and witnessed the dismantling of their entire empire. The more recent horrors of the Thirty Tyrants and the upheaval of democratic forces left the people sensitive to any activities that could further disrupt their already crumbling social and political order. Perceived attacks on religion in times of stress tend to bring out those conservative champions of traditional beliefs who feel compelled to protect their society. One need only look back to the spate of impiety trials during the turmoil of the Peloponnesian war. The reason Meletus and his cohorts targeted Socrates most likely stemmed from his public prominence—and even because of personal animosity. For example, Anytus had personally clashed with Socrates over the career choice of his son (Xenophon,

Apology 29–30). Furthermore, in the antisophist environment of the time, this may have encouraged Meletus and Lyco. Given the animosity that Socrates could provoke and the popular perception of his arrogance, he may have said or done something that antagonized the wrong person at the wrong time.

The last assault on the historicity of Socrates's trial and execution concerns the actual sentence itself. In Plato's *Phaedo*, one is confronted with a Socrates who calmly appropriates the poison. While chastising his friends for their grief, he quips with his executioner and continues to philosophize about the soul. He drank the hemlock, which slowly deadened his legs. He then lay down to wait for the poison to slowly turn his extremities "cold and rigid" (*Phaedo* 118A). Once it reached his heart, he uttered his last words and quietly expired. This scene has been traditionally rejected on the grounds that it does not portray an accurate description of death by poison. Thus, Plato is obviously fictionalizing in order to create a picture of a man with impressive and heroic character, who died in stoic control of his mind and actions (Brickhouse and Smith 2004: 263).

The assaults on Plato's version are commonly found in two recent articles. Christopher Gill commented on the contrast between Plato's depiction of death by poison and that given by other classical writers, such as Nicander. The latter's description is much more violent with convulsions and choking (Gill 1973: 25–28). More recently, Bonita Graves provided more clinical evaluations of hemlock poisoning symptoms in connection with Socrates's death (1991: 156–68). These represent what appeared to be definitive medical proof that Plato had sacrificed historical events for his own philosophical agenda.

Yet, Enid Bloch has impressively dismantled the arguments of Plato's critics (2001: 255–78). She reexamines not only the ancient sources, but also draws exhaustively on historical evidence from medical explorations of the use of hemlock. She effectively dismantles the argument against Plato's veracity. First, she specifies that Plato never actually used the word hemlock, but only "the drug." Later writers identified it with hemlock, which in reality refers not just to one plant, but to a wide variety of plants in the same family, some of which can cause convulsive seizures, but one, which does not (Bloch 2002: 259–60, 262). Second, she provides two accounts that present ways in which the specific hemlock plant is utilized, providing symptoms that coincide with those described by Plato. The 19th-century toxicologist John Harley experimented on himself on two different occasions by taking doses of hemlock (Harley, 1869). In both cases, he described the heaviness of the legs and the onset of paralysis, but commented on his ability to maintain clarity of mind. The second account is the tragic incident of a tailor whose children presented him with what they thought was a wild parsley sandwich. Instead, they accidentally poisoned him. The physician at the hospital interviewed all involved and cataloged the symptoms, which involved a heaviness of the legs, paralysis, and asphyxia with no convulsions

(Bloch 2001: 263–64). In the final analysis, one cannot definitively reject the plausibility of Plato's account of the end of his master's life.

The challenge to the historicity of the events surrounding the death of Socrates will never cease. Indeed, as Guthrie comments, if all scholars agreed on the various issues, then the conclusions about Socrates would be suspect (1969: 8–9). As long as the "Socratic problem" exists, one will always have a seed of doubt about the "historical" Socrates. Yet, certain facts can be maintained plausibly in the face of critics. In 399 BCE, Athens put a man named Socrates on trial for impiety. Nobody in antiquity disputed that fact. The charges are consistent with precedent and seem to be based on actual beliefs about the gods promoted by Socrates (beliefs that neither Plato nor Xenophon needed to fabricate to further their own agendas). That his followers did take liberties with the actual words of Socrates is not in dispute, but once again, the practice is consistent with ancient practice. The actual man himself can still be glimpsed behind the literature of the classical world, and his trial continues to stand as a milestone event in the history of both Greece and humanity.

References and Further Reading

Allen, R. E. *The Dialogues of Plato*, Vol. 1. New Haven, CT: Yale University Press, 1984.

Aristophones. *Clouds*, trans. Jeffrey Henderson in *Aristophanes: Clouds, Wasps, Peace*. Loeb Classical Library, Vol. 488. Cambridge, MA: Harvard University Press, 1998.

Aristotle. *Nicomachean Ethics*, trans. H. Rackham. Loeb Classical Library, Vol. 73. Cambridge, MA: Harvard University Press, 2003.

Bauman, Richard. *Political Trials in Ancient Greece*. New York: Routledge, 1990.

Bloch, Enid. "Hemlock Poisoning and the Death of Socrates: Did Plato Tell the Truth?" in *The Trial and Execution of Socrates: Sources and Controversies*, ed. Thomas C. Brickhouse and Nicholas D. Smith. New York: Oxford University Press, 2001.

Brickhouse, Thomas C., and Nicholas D. Smith., eds. *Routledge Philosophy Guidebook to Plato and the Trial of Socrates*. New York: Routledge, 2004.

Brickhouse, Thomas C., and Nicholas D. Smith., eds. *Socrates on Trial*. New York: Oxford University Press, 1989.

Burnyeat, Myles F. "The Impiety of Socrates," in *Plato's Euthyphro, Apology, and Crito: Critical Essays*, ed. Rachana Kamtekar. Lanham, MD: Rowman and Littlefield, 2004.

Chroust, Anton-Hermann. *Aristotle: New Light on his Life and on Some of his Lost Works*. Notre Dame, IN: University of Notre Dame Press, 1973.

Diogenes Laertius. *On the Lives and Opinions of Eminent Persons in Philosophy*. Translated by R. D. Hicks. Loeb Classical Library, Vol. 184. Cambridge, MA: Harvard University Press, 1925.

Fowler, Harold North. *Plato: Euthyphro, Apology, Crito, Phaedo, Phaedrus*. Loeb Classical Library, Vol. 36. Cambridge, MA: Harvard University Press, 2005.

Gill, Christopher. "The Death of Socrates." *Classical Quarterly* 23 (1973): 25–28.

Graves, Bonita M., et al. "Hemlock Poisoning: Twentieth Century Scientific Light Shed on the Death of Socrates." In *The Philosophy of Socrates*, edited by K. J. Boudouris. Athens, Greece: Kardamista Publishing, 1991.

Guthrie, W. K. C. *Socrates*. New York: Cambridge University Press, 1969.

Harley, John. *The Old Vegetable Narcautics: Hemlock, Opium, Belladonna, and Henbane*. London: Macmillan and Co., 1869.

Kahn, Charles. *Plato and the Socratic Dialogues*. New York: Cambridge University Press, 1996.

Marchant, E. C. *Xenophon: Memorabilia, Oeconomicus, Symposium, Apology*. Loeb Classical Library, Vol. 168. Cambridge, MA: Harvard University Press, 2002.

McPherran, Mark L. *The Religion of Socrates*. University Park: Pennsylvania State University Press, 1998.

Parker, Robert. "The Trial of Socrates: And A Religious Crisis?," In *Religion in Socratic Philosophy*, edited by Nicholas D. Smith and Paul B. Woodruff. New York: Oxford University Press, 2000, 40–54.

Stone, I. F. *The Trial of Socrates*. New York: Anchor Books, 1989.

Stryker, E., and S. R. Slings, "On the Historical Reliability of Plato's Apology." In *Plato's Euthyphro, Apology and Crito: Critical Essays*, edited by Rachana Kametkar. Lanham, Md.: Rowman and Littlefield, 2002.

Thucydides. *History of the Peloponnesian War*. trans. C. F. Smith. Loeb Classical Library, Vol. 108. Cambridge, Mass.: Harvard University Press, 2003.

Vlastos, Gregory. "The Paradox of Socrates." In *Studies in Greek Philosophy: Socrates, Plato, and their Tradition*, Vol. 2, edited by David Graham. Princeton, NJ: Princeton University Press, 1995.

Waterfield, Robin. *Xenophon: Conversations of Socrates*. New York: Penguin, 1990.

Wilson, Emily. *The Death of Socrates*. Cambridge, MA: Harvard University Press, 2007.

CON

The ancient world speculated about the origins of the universe and the natural forces like wind, lightning, and heat in the world. Some of the ancients wrote about these speculations in myths and epics that used events as rhetorical devices rather than actual chronicles of historic events. These myths and epics were testimony that ancient people were always questioning their world as expressed through stories. In this vein, the achievement of the ancient Greeks was that they were willing to deal with these questions in a factual and thoughtful way rather than using stories. They were already asking how the world was made, and they believed that basic elements were the original elements of the world. These ancient people were called the pre-Socratics, and they saw individual facts and created theories to explain the universe through natural laws. Consequently, these pre-Socratics began the elementary roots of science. However, the Greek approach began to change with the rise of a group of itinerant philosophers known as the Sophists. Some philosophers claim that the historic character of Socrates came less from the actual person than from the Sophists stance on the world. The purpose of this section is to trace the origin of Socrates's philosophy and compare it, in terms of style and content, to that of the Sophists' in order to show that the story of the trial of Socrates, as described by Plato, was more likely a sophistic device rather than an actual chronicle of an event.

Pre-Socratics

In discussing Socrates, we need to understand the foundations of philosophy especially with the Greeks. The Greek philosophers asked basic questions about the nature of reality. What is it? What is it made of? What are the elements of world? By what law is nature made? The pre-Socratics, who flourished during the early age of Greek prosperity, thought about these questions. These pre-Socratics, such as Thales and Anaximander, worked on speculating whether water and the boundless or infinite were the foundation of the universe. Heraclitus declared that the fundamental element was fire and that the world had neither beginning nor end. For example, Democritus speculated that atoms were the fundamental units of the universe. These pre-Socratics' thoughts became the fundamentals for the philosophies of the Sophists and Socrates.

The Sophists

The Sophists were a group of teachers who journeyed around Greece to teach young men for money. They questioned the nature of reality and the nature of human conduct. What can we know about nature? Does nature exists outside of human perception? What is the good life? They were instructors of rhetoric, reasoning, and skepticism to the rich people of Athens. Many of the Athenian

---

Socrates on Sophism: Plato's *Protagoras*

Plato reveals much about either his own or Socrates's attitude toward Sophism in his *Protagoras*. The dialogue is between Socrates and Protagoras, the supposed first Sophist:

> You are going to commit your soul to the care of a man whom you call a Sophist. And yet I hardly think that you know what a Sophist is; and if not, then you do not even know to whom you are committing your soul and whether the thing to which you commit yourself be good or evil.
>
> I certainly think that I do know, he replied.
>
> Then tell me, what do you imagine that he is?
>
> I take him to be one who knows wise things, he replied, as his name implies.
>
> And might you not, I said, affirm this of the painter and of the carpenter also: Do not they, too, know wise things? But suppose a person were to ask us: In what are the painters wise? We should answer: In what relates to the making of likenesses, and similarly of other things. And if he were further to ask: What is the wisdom of the Sophist, and what is the manufacture over which he presides?—how should we answer him?
>
> How should we answer him, Socrates? What other answer could there be but that he presides over the art which makes men eloquent?
>
> Yes, I replied, that is very likely true, but not enough; for in the answer a further question is involved: Of what does the Sophist make a man talk eloquently? The player on the lyre may be supposed to make a man talk eloquently about that which he makes him understand, that is about playing the lyre. Is not that true?
>
> Yes.
>
> Then about what does the Sophist make him eloquent? Must not he make him eloquent in that which he understands?
>
> Yes, that may be assumed.
>
> And what is that which the Sophist knows and makes his disciple know?
>
> Indeed, he said, I cannot tell.

Source: Plato, *Protagoras*, Project Gutenberg, www.gutenberg.org/files/1591/1591-h/1591-h.htm. (accessed May 31, 2010)

---

politicians sent their sons to be trained by the Sophists, because the Sophists taught their sons how to reason and argue, which were the basic skills of a politician. Rhetoric was the tool that the Sophists used when attempting to persuade others to their views. They used irony, repetition, and logic to teach their students how to argue with other people. Logic allowed the Sophists to teach the students how to analyze critically the opponent's argument for flaws as well as to organize their speeches to persuade other people with good reasons. They were similar to

Greek red-figure vessel of a Sophist instructing a student in rhetoric, from Cerveteri, an ancient city northwest of Rome, 480 BCE. (Jupiterimages)

the modern-day lawyers of the day, who use the same methods to argue in court for their defendants and questioning witnesses.

In the Sophist questioning, they decided that knowledge was based on individual opinion and perceptions. They decided that a person's experience was relative to him and there were no absolute truths of life. This means that if a person said an object is orange and another person said it was red, they were both right according to their senses. As a result, the Sophists doubted that we can know anything for sure, and all knowledge is relative to the person.

Despite their doubt of knowledge, they believed that the study of human beings and examining their behavior was the proper subject for philosophy. In their education of others, they attempted to understand the best life to lead. They believed that excellence could be taught to their students. This meant that as philosophers they wanted to understand politics, psychology, and government, because these subjects focused on human beings in their teachings. They questioned the customs and traditions of human society because they wanted to know how best to train their students.

The Sophists, due to their questioning of beliefs and through fostering a climate of tolerance for new beliefs, are credited with creating democracy in Athens. Through their training of the leaders of the Athens, they allowed the leaders to learn the process of argumentation and to understand new ideas. These processes allowed the attitude of freedom to be maintained by these leaders. This freedom of the mind was carried into the Athenian government, where through the force of persuasion, they were able to create an open place for the discussion of new ideas. This discussion of opposing arguments allows unpopular views to be voiced in the assembly. As a result, the roots of democracy began with the Sophists teaching these leaders how to reason and to think democratically. However, as Athens grew more powerful, many people wanted a status quo and to keep power in the government. These Sophists would question their leaders' motives and reasons, and through this they created antagonism among the leaders of the day.

Protagoras (490–420 BCE) is generally regarded as the first Sophist. He believed, like Socrates, that virtue could be taught, because men had the ability to learn virtue. Protagoras was the first teacher of virtue, and he studied the use of

language as a way to teach virtue. Protagoras also said that he viewed all truth as dependent on the individual. His famous quote is: "Man was the measure of all things." This statement indicates his belief in relativism. Furthermore, Protagoras believed that we cannot know whether there are gods, because life is too brief and because of the mystery of the subject. However, he viewed that man was the measure of all things or that all truth depended on him. He also said that if you have two arguments, usually the one that you avoid the most is the more truthful of the arguments. This belief shows a type of argument (Guthrie 1971: 262–63).

Others include Gorgias, Prodicus, Hippias, Lycophron, Callicles, and Cratylus. Gorgias (487–376 BCE) originated from Sicily. He made use of paradox and exaggeration as methods of persuasion in rhetoric, which Socrates used in the *Apology* (Kennedy 1972: 31). Prodicus (465–415 BCE) believed in the importance of the exact use of words and style in rhetoric, which Socrates also believed about philosophy. Hippias of Elis (was born about the middle of the fifth century BCE), contributed the meaning of words, rhythm, and literary style to the rhetorical art. Lycophron was a Sophist who believed that law was a convention made by man. Callicles argued that laws were made by men for their own interests and not by gods. Cratylus, another Sophist, believed that words changed so much that communication was impossible. All these Sophists echoed themes that Socrates would emulate or challenge in Plato's dialogues.

Socrates: Background

There was one philosopher thought to be a Sophist, because some of his beliefs were similar. This philosopher was Socrates (470–399 BCE), who was born in the artisan class. He was very educated, and he fought in a war. People thought he was very disciplined, because he could stand for a long time and endure cold weather. He married a woman named Xanthippe, who was his wife for a long time. Socrates used to go into the *agora*, the marketplace, to discuss philosophy with young people. He conducted dialogues discussing ethical and human issues.

Many contemporaries thought he was a Sophist because of the way he conducted arguments. He attempted to conduct a rhetorical discussion like the Sophists would. He would ask questions by giving a definition and answer with another question until the essence of the definition was found. Through constant questioning and doubt, he would find the essence of the words or topic of discussion. He attempted to open the listener to other possibilities than their own beliefs. This was called the Socratic method.

Diogenes Laertius, a Greek historian, wrote the *Lives and Opinions of Eminent Philosophers* in 300 CE. In this book, he claims that Protagoras was the creator of the Socratic method because he created a method of arguing that was similar to the Socratic method. They both argued claims by putting forth a counterclaim and then choosing the one that was least acceptable to argue. By

doing this, they could find the truth by examining the strengths and weaknesses of the counterclaims.

Like the Sophists, Socrates believed that philosophy was the most important study of man. He also believed that excellence could be taught to the student, but that teaching helped the student remember his own thoughts about virtue. Thus, Socrates felt that virtue was an important part of life, and living virtuously was the best way to live like the Sophists.

Socrates had mixed opinions about the Sophists. He did praise them as being better educators than he himself was, and he also sent a student to study under the Sophists. His contemporaries also believed him to be a Sophist, because he taught like them. For example, Aristophanes, a comic playwright, also associates Socrates with being a Sophist.

But Socrates challenged the skepticism and relativism of the Sophists. As a result, Socrates differed with the Sophists on philosophical issues, but he used Sophist tools to seek the truth.

The *Apology*: A Sophists defense

In the *Apology*, Socrates uses many of the Sophists' tools to teach his beliefs about philosophy, virtue, and the moral conduct of life. The charges against him were the same charges that the Athenian leaders leveled against the Sophists. They were questioning the traditions, religions, and philosophical basis of the Athenian society. Socrates was also questioning the basis and the tolerance that the Sophists had developed, because many Athenians were too conservative to allow for tolerance.

Socrates challenged many of the aristocrats as well as taught the children of these aristocrats. Because of these challenges to their thinking, many of the aristocrats and leaders thought he was corrupting the youth of the city and challenging religion. He made these young people think! The leaders felt threatened because this type of thinking could mean the end of their power. Socrates was very similar to the Sophists in that he made people think about the assumptions on which they based their ethics and knowledge. Socrates claimed he had been told by an oracle to embark on this mission of questioning in order to find wisdom. He knew he did not have all the answers, but the leaders thought that they did. Socrates challenged this and took away their pretence to wisdom. As a result of being exposed, they felt threatened because they felt their influence over the people would be reduced. As a result of Socrates's seeking for wisdom, he disturbed the guardians of the society, who brought him up on charges of corrupting the youth. The *Apology* strongly supports the belief that Socrates acted like a Sophist in his life, even up to his death, to the point of using Sophistries to defend himself in the court. Furthermore, Socrates even had people questioning their own immortality and their religion, because he too had to face death because of his Sophist-like attacks against the conservative leaders.

In his use of Sophist tools in the *Apology*, Socrates uses a logical argument of generalization and exceptions in his defense against corruption of youth. Socrates was defending himself against the charges of being a corrupter of youth by comparing himself as only one corrupter of the youth among many supporters of the youth. Socrates raises the question that if there is only one corrupter, how can he influence so many people? Is he the only exception? He is using a Sophistry to logically point out a fallacy or untruthfulness of the charge to his side. This Sophism was displayed when he was questioning Miletus and posing questions like a Sophist would in court.

Socrates was accused in the Athenian court of having taught evil men such as Critias, Alciabiades, and Charmides, who had rebelled and attacked the Athenian state in the past. His defense, another sophistry, was to say that he did not teach them anything or that his teachings had nothing to do with their evil lives. This is a Sophism that attempted to exclude him by explaining that there might be other choices than what they accused him of. He is pointing out that they chose their evil lives and that his teachings had nothing to do with their evil actions.

He adds to this defense that if he had corrupted the youth, he did this involuntarily and so must be absolved of the charges. If a man involuntary commits evil, it may be true in a transcendental way to absolve responsibility. However, in a practical sense, it absolves all criminals of responsibility, because they all would say that they committed it because of other forces. Thus he says that it is relative to the person. He also defends himself by saying the relatives did not charge him with the crime, so why should he be held responsible. These relatives should have witnessed against him for corrupting the youth.

In terms of his defense about not believing in gods, he uses another Sophism by saying that he believes in the sons of gods or lesser divinities. He was also practicing the minimal religious rituals of the times. In saying that he believes in these lesser divinities, he does not commit himself to believing in the state gods or to those of the Greek mythology. He is just saying he does not know about them. This is similar to Protagoras's belief about the gods. In addition, he supports his belief in this demon or guiding spirit that led him to this philosophical journey. Socrates also is saying that this is unimportant to the paramount duty of doing right and self-examination, which is the core of religion.

In the last part of his speech, he tries to distance himself from the Sophists, but maybe he is trying to claim that he is a better Sophist. He believes that he was on a divine mission to seek wisdom and to support the youth, which was accidental and nebulous, but he still went ahead with his questions. He also states that he will subject himself to divine will, because he doesn't know anything about the afterlife, other than as a long sleep, which again cast doubts on issues in which people believe. Socrates also sees that his judges appear a little haughty, but he continues teaching as well as philosophizing until the end in order to get the judges to question their assumptions. In a sense, he was teaching as his

student Plato would teach to a Sophist whom he disliked. However, Socrates attempted to portray human excellence throughout the trial.

Many historians have claimed that Plato hated the Sophists and the Athenians because of their condemnation of Socrates. Because of this hatred, he denied democracy, liberalism, and relativism as threats to true philosophy, which involves finding the ideals of life and the virtues of excellent living. Plato saw that Socrates used many of the Sophists' tools, but he went further in using them to seek the absolute truth and values in life. As a result, Socrates portrayed himself as a transitional philosopher using the Sophists' methods to find the absolute values of the classical philosopher.

References and Further Reading

De Romilly, Jaqueline. *The Great Sophists in Periclean Athens*. Trans. Janet Lloyd. Oxford: Clarendon, 1992.

Denyer, Nicholas, ed. *Protagoras and Plato*. New York: Cambridge University Press, 2008.

Diogenes Laertius. *Lives of Eminent Philosophers*. Trans. R. D. Hicks. 2 vols. Cambridge, MA: Harvard University Press, 1959.

Guthrie, W. K. C. *The Sophists*. New York: Cambridge University Press, 1971.

Kennedy, George. *The Art of Persuasion in Greece*. Princeton, NJ: Princeton University Press, 1963.

Kerferd, G. B. *The Sophistic Movement*. New York: Cambridge University Press, 1981.

Rankin, H. D. *Sophists, Socratics and Cynics*. London: Croom Helm, 1983.

Schiappa, Edward. *Protagoras and Logos*. Columbia: University of South Carolina Press, 1991.

Sprague, Rosamund Kent, ed. *The Older Sophists: A Complete Translation by Several Hands*. Columbia: University of South Carolina Press, 1972.

5

Pushyamitra Sunga, a Hindu ruler in the second century BCE, was a great persecutor of the Buddhists.

PRO   Caleb Simmons
CON   K. T. S. Sarao

PRO

Throughout much of its history, Hinduism has been promoted as a religion of tolerance. Even with the modern movement of Hindu revivalism, where fundamentalism predominates, these same practitioners argue that Hinduism at its core is tolerant of other religious traditions because of the multiplicity of goals and paths covered by the umbrella term *Hinduism*. However, the historical narrative of an individual identified as Emperor Pushyamitra Sunga has surfaced to combat the claims of religious tolerance. Emperor Sunga is said to have persecuted Buddhist monastics. This is a hotly contested issue by Hindus and non-Hindus alike, primarily in India, given the surge for a Hindu nation by the *Hindutva*. Literally "Hindu-ness," the *Hindutva* is a movement that incorporates several political parties that push for the secular democracy of India to be converted to a Hindu state.

To better understand the situation, the character of Pushyamitra Sunga and the text from which we have learned about him must be examined, and the biases the author may have had must be evaluated. Then, the historical setting in which he rose to power must be examined. This setting includes religious, cultural, and political elements. This will show that many of the religious parameters constructed into neatly organized categories simply did not exist in that time. Also, it will become evident that the Buddhists made strong political alliances with several of the enemies of the Sunga Dynasty. Upon the assessment of these elements, it can be concluded that Pushyamitra Sunga did engage in an active campaign against Buddhist monastics, but that the assault was motivated by politics rather than religion.

The Historical Pushyamitra Sunga

Pushyamitra Sunga was born in 215 BCE during the reign of the Mauryan Empire in India. The background of his family is not clearly given in any historical texts. However, in *Brhadaranyaka Upanishad* and the *Ashvalayana Shrauta*

83

Stotra, the Sunga family is referred to as an accomplished group of teachers from the Vishishta-advaita gotra, the school of qualified nondualism started by Ramanuja Achraya. This link would place Pushyamitra within a lineage recognized as one of the six orthodox *darshanas* (philosophical schools) by the *Manavadharmashsta* (The Laws of Manu), a text composed slightly earlier that promoted a social order based on the Vedic society and Brahmanic authority derived from Brahmans, the priestly caste (*varna*) within this system and the center for religious practice and piety.

B. N. Puri has argued that the Pushyamitra is of Persian descent based on the suffix of *mitra* on his name, but this is not likely because the term *mitra* is a common Sanskrit word meaning *friend* or *companion*. Because of the linguistic commonality between Persian and Sanskrit, both being early Indo-European languages, similar terms are not unusual. Therefore, Pushyamitra literally means "the one who is a supreme companion," which, as we will see, is a deceptive title with regard to his relationship with the emperor.

Pushyamitra is found in many historical texts and there is little debate that he was a historical figure. The most important of these texts is the *Mahabhashya*, written by Patañjali. It is unclear if this author is the same as the famous grammarian Patañjali, or if this name was used to legitimize the texts as is sometimes done in ancient cultures. Whatever the case, the *Mahabhashya* contains a detailed history of the rule of the Sunga clan as well as their predecessors, the Mauryas. Despite this account and others like it confirming the existence of a historical character named Pushyamitra Sunga, according to Baij Nath Puri (1957) in *India in the Time of Patañjali*, the *Bhagavata Purana*, a text about the Sunga ruler Bhagavata around 170 CE, makes no mention of Pushyamitra in its listing of the lineage of Sunga rulers. Again the reason for the omission is unclear, but it could be the result of internal struggles for power or an oversight by the sage composing the text, because, as we will see, Pushyamitra had strong ties to the Mauryan Empire and might have been mistaken as the last Mauryan ruler.

Pushyamitra established himself as a great warrior under the Mauryan emperor Brihadratha. Eventually he was promoted to *senapati* (literally, "lord of the army"), the head general over the Mauryan army. According to B. G. Gokhale in *Buddhism and Asoka*, in 188 BCE, while Brihadratha was examining his forces, Pushyamitra Sunga assassinated the emperor and established himself as king. Others have dated his ascension three years later, marking the date of Brihadratha's assassination in 185 BCE. Upon ascending to the throne, Pushyamitra moved the capital of the empire to Pataliputra. He also began to patronize Brahmanic temples and shrines as evidenced in temple inscriptions at Ayodhya, Barganga, and other locations. His militaristic prowess was always upheld, because he continued to refer to himself as *senapati* even after becoming *maharaja* (emperor). These inscriptions also serve as historical markers, because they

were often established after some great feat by the patron. Because of this, however, the accounts are often exaggerated to exalt the patron even more.

Horse Sacrifice Legend

The Ayodhya inscription discusses one very important aspect of the reign of Pushyamitra, the *ashvamedha* horse sacrifice. This story is corroborated in Patanjali's *Mahabhashya* and in popular culture by the second- to third-century CE poet/playwright Kalidasa in his epic drama *Malavakagnimitra*. The *Malavakagnimitra* is centered on Pushyamitra's son and successor, Agnimitra, but goes into detail concerning Pushyamitra's two horse sacrifices. Thus, the tale of Pushyamitra's sacrifices must have been somewhat common knowledge. The horse sacrifice is a very important Vedic ritual in establishing the religious life of Pushyamitra and eventually perceiving his relationship with Buddhism. It was a means by which the social and cosmic orders were upheld and was a form of propitiating the Vedic pantheon. This ritual was only to be performed by "righteous" kings as mandated by *The Laws of Manu*.

The sacrifice also has strong ties to the epic tradition of Brahmanic Hinduism. The Pandavas in the *Mahabharata* and Dasharatha in the *Ramayana*, two of the most important epics of the tradition, perform this sacrifice as the righteous and rightful kings. The sacrifice was used as a threefold confirmation of legitimacy, wealth, and power. The king was legitimated as the rightful king because he was the one who preserved the social order by placating the deities. He displayed his wealth not only through sacrificing his best horse, but also through staging a lavish festival surrounding the ritual. His power was shown by allowing his best horse to wander the kingdom unrestricted for one year. This would demonstrate that as far as the horse could roam in a year was under the protection of the king. This is crucial for the Sunga Dynasty, because both rituals were performed after victorious military campaigns against a Greek coalition, which was formed between the Greeks (*yavanas*)

A horseman, detail from a the gate of a Buddhist stupa, from Bharhut, Madhya Pradesh, Late Mauryan Period, Sunga dynasty, 180–72 BCE. (Indian Museum, Calcutta, India/Giraudon/The Bridgeman Art Library)

and Indians, particularly Buddhists. This will be discussed in much greater detail below.

There may have also been internal reasons for the first of Pushyamitra's *ashvamedhas*. His son, Agnimitra, the central character in Kalidasa's drama, had been sent away to serve as viceroy of Vidisa. Vidisa is also possibly an independent kingdom that Pushyamitra had ruled over while he was *senapati* in the Mauryan kingdom. Given Agnimitra's eventual overthrow of the kingdom, Vidisa might have been where disgruntled generals were sent. There seems to have been a certain tension between the father and son either resulting from this posting or an event that resulted in Agnimitra being sent to Vidisa. The filial ties would soon be mended, because Agnimitra's son, Vasumitra, was chosen to escort the horse during its year of roaming. This was a high honor usually bestowed on the kingdom's greatest warrior. This act would solidify the relationship and secure the stability of the Sunga Dynasty and its succession.

Persecution of Buddhist Monastic Community

Thus far there has been no mention of persecution of the Buddhist monastic community, although it is evident that Pushyamitra has strong ties to Brahmanic Hinduism. The ancient sources that accuse Pushyamitra of slaughtering Buddhists are the *Divyavadana*, a Buddhist text that contains a history of Indian rulers and their relationship to the *sangha* (the Buddhist community) and the *Ashokavadana*, the narrative of the life of Ashoka Maurya. In a very brief account, the *Divyavadana* states that Pushyamitra Sunga was one of the greatest persecutors of the Buddhist community in Indian history. According to this account, Pushyamitra sought council from various ministers as to how to become as famous as the great Ashoka, a Mauryan ruler who converted to Buddhism, which will be discussed in greater extent below.

Several ministers explained that Ashoka had become so popular by establishing 84,000 Buddhist reliquaries (*stupas*) around the empire and suggested that he do the same. However, a Brahman minister suggested that he would become more famous than Ashoka if he were to eliminate Buddhism and the *stupas*, and thus the memory of Ashoka. Pushyamitra heeded the Brahman's advice and began attacking Buddhist monasteries. Pushyamitra not only attacked the monks and nuns, he is also reported to have offered a reward for anyone who killed a monk upon the presentation of the dead body. The account given in the *Ashokavadana* is quite similar but goes into greater detail on Pushyamitra's use of the four-element army consisting of elephants, cavalry, chariots, and infantry in his attack on Buddhist monasteries. This presentation of Pushyamitra Sunga is quite different from most of the previous examples given. Therefore, the context of Pushyamitra's ascension and the shift in religious patronage must be examined to reconcile the various accounts.

Jain Dharma

Focusing on self-discipline, nonviolence, and education, Jainism or Jain Dharma originated in India in the sixth century BCE. It shares with Buddhism and Hinduism the notion of karma and the quest for enlightenment; like some sects of both of those faiths, the reverence for life leads to the practice of vegetarianism, veganism, and even the sweeping of the ground before one's footsteps so as to avoid killing insects underfoot. The Jain cosmology is similar to that of Hinduism, and the Ramayana and Mahabharata are important texts, though Jains interpret them somewhat differently. There is also a similarity to many Christian and Jewish sects in antiquity: fasting and a denial of material wealth and worldly goods are important aspects of many Jain practices. Jain monks of the Digambar sect do not even wear clothes, considering them an unnecessary material possession.

Rulers Prior to Sungan Rule

Prior to the Sungan rule, the majority of the Indian subcontinent had been ruled by the Mauryan Empire. Since its inception, the dynasty had been very diverse in its religious patronage. The founder of the empire, Chandragupta Maurya (ruled 322–298 BCE) was a patron of the Jain tradition established by Mahavira. This tradition was considered a heretical *shramana* movement by followers of the Brahmanic tradition. The *shramanas*, which also include Buddhism, were traditions that denied the authority of the Vedas and the caste system. These teachings placed the *shramana* traditions at odds with Brahmanic tradition because both the Vedas and caste were their central tenets of belief. Chandragupta, however, also patronized various temples that would now be classified as Hindu. This tradition of patronage to several varying religious tradition was not unusual for the time.

The neatly divided categories of "Hindu," "Buddhist," or "Jain" did not exist yet. Therefore, the ability to follow a multiplicity of paths was not uncommon. The dominant religious tradition was what has been described above as Brahmanic Hinduism, but to view this as a unified movement would be anachronistic. The beliefs and ritual would vary widely based on region and cultural differences. The common thread was the Vedic sacrifice and caste, with the priest class as the sacrificial officiates, but different shrines that followed a different philosophical path (*darshana*) or were devoted to a different deity would jockey for position even within what today we would call Hinduism. Thus it would be a fallacy to view the positioning for patronage as a battle simply between Hindus and Buddhists. It was a much more complex arena. However, with the conversion of Emperor Ashoka, the tension between the loosely connected Brahmanic traditions and the Buddhists grew and perhaps united the groups with another thread, that of anti-*shramana*.

Chandragupta is also important to our discussion because of his relationship with the Greek Seleucids. According to Appian of Alexandra in his *History of Rome: The Syrian Wars*, when Seleucus I entered India to conquer the Mauryan territory, he was met with greater resistance than he had planned. Thus he entered into a marriage alliance with Chandragupta. In accordance with the treaty, Seleucus was also given 500 war elephants and Chandragupta received several Persian provinces. Several Greek historians also remained in the Mauryan court to chronicle the new Greco-Indian ally.

Ashoka Maurya

Ashoka Maurya (ruled 273–32 BCE) was the third ruler of the Mauryan Empire. Early in his reign, he conducted many great military campaigns and eventually united all of south Asia under his banner. However, after his last victory in Kalinga, he realized the suffering caused by his brutal search for power and converted to Buddhism. He then began promoting the Buddhist doctrine of non-violence (*ahimsa*). This religious experience was also an interesting political move that would discourage violent coups d'état of his government, but would also alienate many followers of the Brahmanic traditions.

Ashoka began to promote Buddhism in many ways that made it the dominant religious tradition of the time. In addition to the 84,000 *stupas* he was said to have constructed, he sent missionaries throughout the empire and beyond. He even sent his son Mahindra and his daughter Sanghamitra, both Buddhist monastics, to the island of Ceylon (Sri Lanka) to spread the Buddhist Dharma. He himself is said to have developed a lifestyle very close to that of an ascetic. Much of the history of Ashoka's promotion of Buddhism is preserved in rock edicts he constructed throughout India.

Many of these edicts suggested religious tolerance; however, his patronage was causing Buddhist monasteries to grow while many of the Brahmanic temples and shrines suffered from the lack of support. The patronage of the Buddhist monasteries was so great, according to Hirakawa Akira's (1990) *A History of Indian Buddhism*, that it began to stifle the economy. Patronage was not the only problem that arose from Ashoka's promotion of Buddhism. In accordance with the doctrine of *ahimsa*, Ashoka made animal sacrifice, the center of Vedic ritual, illegal. Therefore, he essentially outlawed the practice of the Brahmanic tradition. As a result, a Brahmanic countermovement was established, with the *Dharmashastra*s (Law Codes) as the collections of central rules by which life should be conducted, centered on sacrifice and purity and pollution of caste. Pushyamitra later became the epitome of this countermovement.

Ashoka also remained allied with the Greeks. He continued to exchange ambassadors with them and even mentioned the Greek rulers in many of his rock edicts. After Ashoka, the Mauryan Empire continued this relationship with

Buddhists and the Greeks, but gradually the rulers became weaker and weaker until Pushyamitra successfully overthrew the empire by assassinating Brihadratha in front of the army. After the assassination, the Mauryan Empire became divided. The northwestern portion was subsumed into the Greek Bactrian Empire, the central region became the Sunga Empire, and the eastern region was reclaimed by the Kalingas, led by the Jain king Kharavela. Thus the majority of the subcontinent abandoned patronage and practice of the Buddhist tradition. Because of the loss of much patronage by the Sunga Empire, tension arose between the Buddhists and the rulers. With the Greeks nearby and former generals, such as Yajnasena, forming hostile relationships with the Sungas and establishing their own independent states, the stage was set for the Buddhist monasteries to become political adversaries of the Sunga Empire.

The Pushyamitra Sunga Era

Pushyamitra Sunga began to distance himself from the practices of the Mauryan rulers and to establish a lineage that would place himself as the rightful king. This practice can be seen throughout south Asia when new rulers rose to power. The change was not regarded as new and innovative, but as potentially detrimental; therefore, the emperor must have been positioned within the broader history of the region. The horse sacrifice also served as a form of legitimization. The first of the two horse sacrifices was performed in Pataliputra, the very site of the edict where Ashoka had forbidden animal sacrifice. It also served as the substitute for the festivals called *samayas* that the Mauryan Empire had established. Thus Pushyamitra was seeking not only authentication, but also the approval of the people. His need for authentication by both people and the scribes that formed lineages through ancestry charts seems to validate the conversation that was the setting for the mention of Pushyamitra in the *Divyavadana*.

While Pushyamitra seems to have been accepted by many of the people, there was a movement against his rule. Puri suggests that a rival Buddhist viceroy from the then-defunct Mauryan Empire, Damstranivasin, who had established an independent kingdom, offered the Greek king Demetrius his daughter if he attacked the Sungas. P. C. Bagci (1946), in an *Indian Historical Quarterly* article, suggests that the Krmisa mentioned in *Manjusrimulakalpa* is Demetrius. The narrative seems to fit, but his conclusion is not accepted by all scholars. Damstranivasin beseeches Demetrius for the alliance on the basis that Pushyamitra is anti-Buddhist. Since the Sunga Empire had usurped the power of a Greek ally, Demetrius agreed, and the first of two Greek invasions during the reign of Pushyamitra had begun. Demetrius is also linked to the Buddhists in the Hathigumpha Inscription, where he is referred to by the name Dharmamitra, the friend of the Dharma. These invasions are widely documented in both Indian and Greek

chronicles as well as archaeological evidences. The motivation, whether political or religious, seems to have been steeped in the conflict of Buddhism and Brahmanism.

First Greek Invasion of Northern India

During this time of turmoil concerning the power structure of the former Mauryan Empire, the Greeks under Demetrius, whether because of an alliance with Damstranivasin or simply to expand their kingdom, invaded northern India. The exact date of the invasion is unclear, but most scholars would agree that the *Yavana* invasion of Demetrius came within the first 10 years after Pushyamitra claimed the throne of Magdha by killing Brihadratha, ca. 187–77 BCE. Details of the invasion are given in several sources, including Kalidasa and Patañjali, as well as in the *Gargi Samhita* in the *Yuga Purana* and in an inscription at Hathigumpha. In *Buddhism and Asoka*, B. G. Gokhale notes that Greek coins featuring Demetrius wearing the scalp of an elephant, an obvious link to India, have been found well into the central region of Sunga territory.

The invasion was very successful in the beginning. The forces led by Demetrius conquered the lands of northern India, forcing Pushyamitra Sunga to retreat and move his capital to Rajagriha, where he would subsequently be attacked by the Kalinga king Kharavela, who obviously felt threatened by the approaching army. Amid the attacks mounting from both sides, Pushyamitra had a brush of good fortune. Demetrius and his forces learned of a civil war that was breaking out in his kingdom, and he was forced to withdraw from the campaign. Pushyamitra seized the opportunity to present himself as the *senapati* who had repelled the foreign invasion. It was at this moment in history that Pushyamitra offered the first of the horse sacrifices in Pataliputra as an example of his power and in no doubt to repay the fortunate turn of events that the gods had orchestrated. Thus Pushyamitra Sunga was able to solidify himself as the emperor of northern India.

Rule of Pushyamitra in Northern India

With Pushyamitra now the uncontested ruler, he began to punish those who had sided against him. Among the most outspoken of these groups were the Buddhist monasteries. The Buddhists in the western Punjab region had openly sided with the Greek invaders, so they were treated as political enemies of the state. Had Pushyamitra been lenient on the traitors, he would have been viewed as a weak ruler, which would have facilitated attempts to overthrow him. In this context, the claims of Buddhist texts alleging cruelty against the Buddhist monasteries and a bounty for the heads of dead monks during Pushyamitra's reign seem fully plausible. His attempt to extinguish the Buddhist *sangha* correlates with his attempts to extinguish political rivals. In the *History of the Sunga Dynasty*, B. C. Sinha (1977) holds that later in his reign, Pushyamitra lessened his persecution of the *sangha*, and the construction of several Buddhist *stupas*

are attributed to his patronage, although this has been disputed by several scholars. However, this leniency was short lived because the second Greek invasion also was tied to the Buddhist community.

Second Greek Invasion

The second Greek invasion during the rule of Pushyamitra came near the end of his reign. The leader of this invasion was the Greek king Menander, Demetrius's successor, who is forever linked with Buddhism as the pious King Milinda of the *Milinda Panha*. His Bactrian kingdom extended from Kabul in modern-day Afghanistan to the Punjab in modern-day Pakistan, where his capital of Sagala was located. He is said to have converted to Buddhism after his discussion with the Buddhist sage Nagasena, as told in the *Milinda Panha*. From Menander's conversion onward, he was a great patron of Buddhism and its monasteries within the subcontinent's northwestern regions.

Much of the Buddhist art of the time reflected Greek influence. In iconography, the Buddha is portrayed with curled hair, a popular characteristic of Greek gods and busts. Even representations of Menander have elements of Buddhism within them. In a relief of the king found on the *stupa* in Bharhut, Menander is depicted holding an olive branch, which shows his relationship to Greek goddess Dionysus, and a sword with the symbol of the three jewels of Buddhism. His coins also illustrate the importance of Buddhism to his court through the depiction of the Buddhist eight-spoked wheel. Thus Buddhism is linked to one of the Sunga's neighboring kingdoms against which they vied for territory and power.

The details of the second Greek invasion are slightly more nuanced than the campaign of Demetrius. The motivation behind the campaign is not entirely obvious. What is clear, however, is that the conflict took place while the horse for the second sacrifice was roaming freely under the guard of Pushyamitra's grandson Vasumitra and his forces. It seems as though Menander took this ritual as an opportunity to attack the divided forces of Pushyamitra. Sinha argues that the horse might have wandered past the Indus River (*Sindu*) into the Greek territory. This would have been an affront to the Bactrian king for several reasons. First, if the horse were to roam into the terrain of Menander, the Sunga legion would follow it. The presence of a foreign army would certainly be enough to incite conflict. Also, the presence of the horse could be taken by Menander as a claim by Pushyamitra to the region, because the purpose of the ritual was to show the far reaches of his kingly influence. The theory of the horse breaking the boundary of the Indus seems less likely, because evidence shows that Menander's armies marched into the middle of the country (*madhyadesa*) before being pushed back and ultimately defeated in a decisive battle at the banks of the Indus by the legion under Vasumitra that was guarding the sacrifice horse.

The second invasion proved to be the flagship by which the Sunga Empire was solidified. The Greek armies were quickly expelled by the Sungan forces, and the war against Menander was exalted as a total annihilation of the enemy. However, the complete vanquishing of the Greek forces must be called into question, because Menander was able to expand his territory in the western portion of the Aryavarta, which, given its name (literally, "the Noble Sustenance"), must have been a region that produced great wealth for the ruler. Despite this small setback, Pushyamitra staged the more elaborate of the *ashvamedhas* and established himself and the Sunga Empire for several generations. With a stronghold on the region for many years to come, Pushyamitra's successors were able to loosen their policies concerning Buddhists and once again patronize multiple religious traditions.

Conclusion

In conclusion, Pushyamitra Sunga actively engaged in a war against Buddhism with many different motivations, most of which were directed back to the establishment of political stability for himself and his lineage. Coming as he did from an orthodox Brahmanic family, Pushyamitra was aware of the Mauryan preference for the anti-Brahmanic *shramana* movements. He may have even been directly affected by the lack of patronage given to Brahmanic pandits if his connection to the Vishishta lineage is accurate. Upon usurping the throne, he needed to create a niche for himself in light of the great Emperor Ashoka; therefore, he distanced himself in religious patronage and sought to erase many of the injunctions created by the former king as retold in the *Divyavadana* and the *Ashokavadana*. The conflict was then multiplied when Buddhist monasteries sided with the foreign Greek invaders led by Demetrius. As a result, Pushyamitra performed an animal sacrifice that directly transgressed Buddhist doctrine. Then he sought to eradicate the traitors by attacking them with his armies and offering rewards for their slaughter. After a period of brief relaxing of his sanctions, a conflict arose between the Greco-Buddhist king Menander and the Sungas. While the policies of Pushyamitra were harsh, they were clearly not misguided. He saw Buddhism as a threat to his ascension to the throne of the empire, and this proved to be true several times. It was only through the elimination of many Buddhists and their allies that the Sunga Dynasty was established and found stability in the region.

References and Further Reading

Akira, Hirakawa. *A History of Indian Buddhism: From Sakyamuni to Early Mahayana*. Honolulu: University of Hawaii Press, 1990.

Albright W. F., and P. E. Dumont. "A Parallel between Indic and Babylonian Sacrificial Ritual." *Journal of the American Oriental Society* 54, 2 (June 1934): 107–28.

Bagci, P. C. "Krmisa and Demetrius." *Indian Historical Quarterly* 22 (1946).

Bhandarkar, Ramkrishna Gopal. *A Peep into the Early History of India*. Bombay: D. B. Taraporevala, 1930.

Cowell, Edward B., and Robert A. Neil. *The Divyavadana: A Collection of Early Buddhist Legends*. Amsterdam: Oriental Press, 1970.

Gokhale, B. G. *Buddhism and Asoka*. Baroda, India: Padmaja Publications, N.D.

Nath, Jagan. "The Pusyamitras of the Bhitari Pillar Inscription." *Indian Historical Quarterly* 22 (1946).

Puri, Baij Nath. *India in the Time of Patanjali*. Bombay: Bhartiya Vidya Bhavan, 1957.

Sinha, Binod Chandra. *History of the Sunga Dynasty*. Varanasi, India: Bharatiya Publishing House, 1977.

Smith, R. Morton. "On the Ancient Chronology of India (II)." *Journal of the American Oriental Society* 77 (October–December 1957): 266–80.

Smith, R. Morton. "On the Ancient Chronology of India (III)." *Journal of the American Oriental Society* 78, 3 (July–September 1958): 174–92.

Strong, John S. *The Legend of King Asoka, A Study and Translation of the* Asokavadana. Princeton, NJ: Princeton Library of Asian Translations, 1983.

Upreti, Kalpana. *India as Reflected in the Diyavadana*. New Delhi: Munshriram Manoharlal Publishers, 1995.

CON

King Pushyamitra Sunga (ca. 184–48 BCE) is generally regarded as the symbol and leader of the Brahmanic revival that took place when the dynasty of the Mauryan kings, the alleged supporters of non-Brahmanic faiths, was brought to an end. The majority of the textual sources dealing with the Sungas link them to the Brahman caste. Thus, the end of the Mauryan Dynasty at the hands of Pushyamitra Sunga is seen as a victory of Brahmanic anti-Buddhist forces that had been silently at work. In other words, it is generally held that after the end of the Mauryan rule, Buddhism not only lost the royal favors it had enjoyed under kings such as Ashoka, but, as a result of the persecution by Pushyamitra Sunga, it also lost most of what it had gained earlier. Thus it has been suggested that other than destroying Buddhist monasteries and *stupas* and killing Buddhist monks, he caused greater damage to Buddhism by letting unfavorable forces loose against it. As we will see, nothing could be further from the truth.

Pushyamitra Sunga was the commander in chief of the last Mauryan king, Brihadratha. After assassinating his master, he captured power and laid the foundations of the Sunga Dynasty. His extensive empire, with its capital at

---

Ashoka (304–232 BCE)

One of the last Mauryan rulers, Ashoka governed the Indian empire when it was at its largest: his territory stretched from Afghanistan and Iran to Assam and Bangladesh and covered most of the Indian subcontinent. His military prowess helped to expand the previous bounds of the empire, and at some point in his 40-year reign he became a convert to Buddhism, which was then in its early centuries. Ashoka's conversion is traditionally credited to guilt over his treatment of his brothers (whom he executed) and the death toll of his victory in the Kalinga War. Whatever the case, his piety was noted by his people, and he left behind 33 inscriptions in stone—on pillars, boulders, and in caverns—popularly called the Edicts of Ashoka.

The edicts are mostly moralist and prescriptive, calling for kindness, generosity, fairness in the execution of justice, and the protection of many animals (including, pragmatically, all four-footed animals that are not useful or edible). It is a feature of the edicts that they generally call for behavior that is moral in intent but practical in execution, not focused too greatly on self-denial or self-burden.

Pataliputra (modern Patna), included the cities of Ayodhya, Vidisha, and Vidarbha (Berar) and extended in the south up to the Narmada River. The accounts in the *Ashokavadana*, the *Divyavadana*, and *Tāranātha's History of Buddhism* also show him as being in control of Jalandhar and Shakala in the Punjab.

Warfare seems to have been the mainstay of the reign of Pushyamitra and other kings of his dynasty. He and his descendants warred with the Andhras, Kalingas, Greco-Bactrians, and possibly the kingdoms of Panchala and Mathura (which may not have been under Pushyamitra's rule). Pushyamitra himself fought at least three major wars. One of these wars was fought against Yajnasena, the king of Vidarbha, who had remained loyal to the Mauryan Dynasty after the coup d'état. He fought the other two wars against the Greco-Bactrians, in all probability against King Menander (Milinda of the Buddhist text *Milindapanha*), whose kingdom lay to the northwest of his empire. The Greco-Bactrians had attacked northwestern India in circa 180 BCE. They eventually captured much of the Punjab, perhaps ruling from Mathura for a while, and may have even ventured as far as Pataliputra. However, in the end, Pushyamitra's forces in all likelihood recaptured Mathura toward the close of the second century BCE and may have driven the Greco-Bactrians out of the Punjab.

After establishing himself firmly on the throne, according to N. N. Ghosh (1945), Pushyamitra is alleged to have run the affairs of his kingdom with the help of his contemporary Brahman scholars, such as Manu (the author of the *Manusmriti*) and Patañjali (the author of the *Mahabhashya*) and reestablished the sacrificial ceremonies of Vedic Brahmanism. As animal sacrifices and old Vedic rituals

were completely discouraged by the Mauryan rulers, who were followers of heterodox faiths such as Buddhism and Jainism, his performance of two *ashvamedha yajnas*, as K. P. Jayaswal (1918) writes, is viewed as an anti-Buddhist activity of a king who was a fundamentalist Brahman. Scholars such as Haraprasad Sastri (1910) feel that actions such as discouraging the animal sacrifices by Ashoka were a direct attack on Brahmans, because much of their power and prestige lay in the fact that they alone could perform sacrifices and thus act as intermediaries between the people and the gods. Such actions, these scholars tell us, deprived Brahmans of their long-enjoyed privilege of guiding the religion of the masses.

Sastri further maintains that the *dhamma-mahamattas* (ministers of piety) employed by Ashoka for the propagation of his policies destroyed the reputation of the Brahmans and that such an action was particularly resented by the Brahmans, because it came from a *shudra* king. Further, it is alleged, Ashoka had acted against Brahmanism by "showing up the false gods," who, until then, had been worshiped in Jambudvipa. Another well-known scholar, U. N. Ghoshal (1966), also feels that the propagation of Buddhism during the Mauryan period had disturbed the Brahmanic social and religious order.

There are also some scholars who feel that even if the atrocities committed by Pushyamitra as reflected in the *Divyavadana* are viewed as exaggerated, the acute hostility and tensions between Pushyamitra and the monks could not be denied. There thus are many scholars who view King Pushyamitra Sunga as a fanatical Brahman king who persecuted and tyrannized the Buddhists by killing them and destroying their holy places. Then there are some other scholars who take a somewhat lenient view and believe that Pushyamitra may not have been an enemy of Buddhism but that he certainly withdrew royal support and was responsible for creating an environment that the Buddhists found unfavorable for their own activities.

Accounts of Pushyamitra's So-Called Anti-Buddhist Activities

To critically examine the various textual references that mention the so-called anti-Buddhist activities of Pushyamitra Sunga would be worthwhile. The most important and perhaps earliest reference is from the *Divyavadana* (and its constituent part, the *Ashokavadana*). According to this Sarvastivadin text of the second century CE, when Pushyamitra asked his ministers how he could obtain everlasting fame, most of them told him that, as long as Buddhist law remained, he would have to construct 84,000 *stupas* as his predecessor King Ashoka had, but one of the priests told him that he could obtain everlasting fame by doing the opposite—destroying the Buddhist religion. According to this text, Pushyamitra chose the latter route and also put a bounty on the heads of Buddhist monks. Then he went to the Kukkutarama monastery at Pataliputra, intending to destroy the Buddhist region. According to these ancient accounts, he was turned away three times at the gate of the monastery by a lion's roar before he

slaughtered the monks and destroyed the residence of the organization. Continuing in this way through the country, he arrived in Shakala, the modern-day Sialkot in the Pakistani Punjab, where he issued the edict of awarding a gold piece for each head of a Buddhist monk that was brought to him. According to the legend, this activity was only ended when the Yaksha Damshtranivasin, the guardian spirit of Bodhi living in this region, enabled Pushyamitra's army to be crushed, Pushyamitra to be killed, and the Maurya dynasty to come to an end.

The *Vibhasha*, a Sarvastivadin-Vaibhashika text dated in the second century CE, chronicled that Pushyamitra hated the Buddhist religion and burned holy books; destroyed *stupas*; demolished monasteries, including 500 monasteries on the borders of the kingdom of Kashmir; and slaughtered monks. According to this source, Pushyamitra was supported with *kumbhandas, yakshas*, and demons that enhanced his powers and made him invincible until he approached the Bodhi tree during his destruction of Buddhism and was vanquished by the deity of that tree, which had taken the form of an extremely beautiful woman in order to be able to approach the king.

This story is also repeated in the *Shariputraparipriccha*, a Mahasamghika text translated into Chinese between 317 and 420 CE. But the story in this text, besides being much more detailed, shifts the anti-Buddhist operations of Pushyamitra Sunga from the northwestern part of the Indian subcontinent to Bihar in the east,

The *Aryamanjusrimulakalpa*, which belongs to the early medieval times, mentions Pushyamitra Sunga in abusive terms such as *Gomimukhya* (cattle-faced) and *Gomishanda* (Gomin, the bull) in an allusion to the Vedic sacrifices that were revived under the Sungas. This text, while talking about the "evil actions" of Pushyamitra against Buddhism, tells the legend of Pushyamitra demolishing monasteries and venerable relics and killing monks after capturing the east as well as the entry into Kashmir and being defeated and dying after he turned north.

Also, Tāranātha, the celebrated Tibetan Buddhist historian, mentions that "the *brahmana* king Pushyamitra, along with other *tirthikasās*, started war and thus burnt down numerous Buddhist monasteries from the *madhyadesha* to Jalandhara. They also killed a number of vastly learned monks. As a result, within five years, the doctrine was extinct in the north" (Tāranātha 1970: 121).

Archaeological Evidence Supporting Persecution Hypothesis

Besides the textual evidence, archaeological evidence is also put forward in support of "anti-Buddhist" actions of Pushyamitra and other kings of the Sunga Dynasty. For instance, John Marshall (1955, 1975) writes that at Takshashila there is evidence of some damage done to the Buddhist establishments about the time of the Sunga. Marshall proposes that the Sanchi *stupa* was vandalized during the second century BCE before it was rebuilt later on a larger scale, suggesting the possibility that the original brick *stupa* built by Ashoka was

Buddhist stupa at Sanchi, India. (Mark Weiss)

destroyed by Pushyamitra and then restored by his successor, Agnimitra. Similarly, N. N. Ghosh (1945) writes that the gateway of Bharhut was built not during the reign of Pushyamitra but by his successors, who followed a more tolerant policy toward Buddhism as compared to Pushyamitra, a leader of Brahmanic reaction. The destruction and burning of the great monastery of Ghositarama at Kaushambi in the second century BCE is also attributed to the Sungas. For instance, J. S. Negi (1958) notes that G. R. Sharma, who was responsible for most of the excavation work at Kaushambi, was inclined to connect this phenomenon with the persecution of Buddhism by Pushyamitra. Similarly, according to P. K. Mishra:

> Although archaeological evidence is meager in this regard, it seems likely that the Deorkothar stupa, geographically located between Sanchi and Bharhut, was destroyed as a result of Pushyamitra Sunga's fanaticism. The exposed remains at Deorkothar bear evidence of deliberate destruction datable to his reign. The three-tiered railing is damaged; railing pillars lie broken to smithereens on stone flooring. Twenty pieces of pillar have been recovered, each fragment itself fractured. The site offers no indication of natural destruction. (2001)

Critique

Many Indologists, including K. P. Jayaswal (1923), H. C. Raychaudhary (1923), R. C. Mitra (1954), and D. Devahuti (1998), have expressed skepticism about

the truthfulness of the Buddhist legends regarding the persecution of Buddhism by Pushyamitra Sunga. Raising serious doubts about the authenticity of the legend, Etienne Lamotte (1988), for instance, has pointed out that the only point on which the sources concur is the destruction of the Kukkutarama of Pataliputra "in the east." If there was an encounter between Pushyamitra and the Yaksha Damshtranivasin and Krimisha, it is impossible to pinpoint where exactly it took place: at Sthulakoshthaka in the Swat Valley, at the Dakshinavihara on the heights above Rajagriha or in Avanti, at the gates of Kashmir, or in Jalandhar. This is also the case with the death of Pushyamitra, which variously takes place under the Bodhi tree at Bodh-Gaya, on the shores of the southern ocean, or somewhere "in the north." Thus Lamotte points out, to judge from the documents, Pushyamitra "must be acquitted through lack of proof" (1988: 109). Agreeing with Lamotte, D. Devahuti (1998) also feels that the account of Pushyamitra's sudden destruction with all his army, after his promulgation at Shakala of a law promising *dinaras* for the heads of Buddhist monks slain by his subjects, "is manifestly false." Taking recourse to similar argument, R. C. Mitra, too, feels that "The tales of persecution by Pushyamitra as recorded in the *Divyavadana* and by Taranatha bear marks of evident absurdity" (Mitra 1954: 125).

H. C. Raychaudhury (1923) and Romila Thapar (1991) also do not believe in the persecution theory. Raychaudhury, for instance, points out that the ban on animal sacrifices did not necessarily entail antagonism toward the Brahmans for the simple reason that the Brahmanic literature itself lays stress on *ahimsa*. For instance, the *Chandogya Upanishad* mentions the importance of nonviolence and the futility of giving too much importance to sacrifices alone. Ashoka did not only ban the sacrifice of those animals that were sacrificed in *yajnas*, but even others.

Thapar opines that Ashoka's frequent exhortations in his edicts for showing due respect to Brahmanas and Shramanas hardly point to his being anti-Brahmanic in outlook. In fact, Raychaudhury notes that some of the *dhamma-maha-mattas* were concerned specifically with safeguarding the rights and welfare of the Brahmans. Haraprasad Sastri's (1910) contention is that Ashoka was powerful enough to keep the Brahmans under control, but after him a conflict began between his successors and the Brahmans that only ended when Pushyamitra assumed power, and that Pushyamitra's action was the manifestation of a great Brahmanic revolution is also indefensible.

As pointed out by Raychaudhury, some of the Mauryan kings were themselves followers of Brahmanism. For instance, according to the *Rajatarangini*, a text belonging to the early Medieval period that deals with the history of Kashmir, Jalauka was not only a zealous *Shaiva* and an open supporter of Brahmanism, but he was also quite unfriendly toward Buddhism. And Thapar writes that "since the Mauryan empire had shrunk considerably and the kings of the later period were hardly in a position to defend themselves, it did not need a

revolution to depose Brihadratha" (Thapar 1991: 201). The fact that Pushyamitra was assassinated while he was reviewing the army does not indicate there was a great revolution. On the contrary, it points rather strongly to a palace coup d'état taking place because by this time the organization of the state had sunk so low that subordinate officials were happy to work under anyone who could give them assurance of a more competent administration. Moreover, as Thapar points out, had it been a great Brahmanic revolution, Pushyamitra would have received the assistance of other neighboring kings such as the descendants of Subhagasena from the northwest.

The testimony of the Buddhist legends also appears doubtful on various other counts. The earliest texts that mention these legends are chronologically far removed from the Sungas. The traditional narrative in the *Divyavadana*, for instance, can, at the earliest, be dated to two centuries after Pushyamitra's death. It is more likely that the *Divyavadana* legend is a Buddhist version of Pushyamitra's attack on the Mauryas and reflects the fact that, with the declining influence of Buddhism at the imperial court, Buddhist monuments and institutions would naturally receive less royal attention. Moreover, the source itself, in this instance being Buddhist, would naturally exaggerate the wickedness of anti-Buddhists.

Further, at the time of the Sungas, *dinara* coins (Roman *denarious* gold coins) were not prevalent. The earliest period during which they came into circulation in India was the first century CE. Most interestingly, this legend of persecution in which a *dinara* is offered as an award for the head of a monk is first related in the *Ashokavadana* in connection with the persecution of the Jainas and the Ajivikas by Ashoka and clearly appears to be a fabrication. To say that Ashoka, whose devotion to all religious sects is undeniable through his edicts, persecuted the Nirgranthas or the Ajivikas is simply absurd, and so is the story of Pushyamitra Sunga. Thus, as pointed out by Koenraad Elst, "the carbon-copy allegation against Pushyamitra may very reasonably be dismissed as sectarian propaganda" (Elst, 2005).

Probity of the *Divyavadana* is also grievously marred by the fact that Pushyamitra Sunga is mentioned as a descendant of Ashoka, whereas he did not belong to the Mauryan Dynasty, a dynasty of non-Brahman background. In fact, this very fact flies in the face of the hypothesis that Pushyamitra persecuted the Buddhists because he was a Brahman. Similarly the argument that the Brahman backlash became intense because the Mauryas were *shudras* does not seem to hold. Besides the fact that the Mauryas are mentioned as *kshatriyas* in the *Divyavadana*, Raychaudhury has pointed out that the *Purana* statement that all kings succeeding Mahapadma Nanda will be of *shudra* origin implies that Nanda kings after Mahapadma were *shudras* and not the Mauryas, because if it referred to succeeding dynasties, then even the Sungas and Kanvas would have had to have been listed as *shudras*.

There is really no concrete evidence to show that any of the Maurya kings discriminated against Brahmanism. Ashoka, the most popular Maurya king, did

not appear to have any vulgar ambition of exalting his own religion "by show-ing up the false gods" of Brahmanism. Thus the theory of a Brahmanic persecu-tion under Pushyamitra loses much of its raison d'être. The policy of Pushyamitra Sunga appears to have been tolerant enough for the simple reason that if he were against the Buddhists, he would have dismissed his Buddhist ministers. Furthermore, the court of Pushyamitra's son was graced by Bhagavati Kaushiki, a Buddhist nun. Moreover, there is overwhelming evidence to show that Buddhism actually prospered during the reign of the Sunga kings.

As Thapar notes, many scholars have actually argued that archaeological evidence casts doubt on the claims of Buddhist texts of persecution by the Sun-gas. An archaeological study, recorded in *Ancient India,* of the celebrated *stupa* at Sanchi proves that it was enlarged and encased in its present covering during the Sunga period. The Ashokan pillar near it appears to have been willfully destroyed, but Marshall (1955) suggests that this event may have occurred at a much later date. According to D. C. Sircar (1965), the Bharhut Buddhist Pillar Inscription of the time of the Sungas actually records some additions to the Buddhist monuments "during the supremacy of the Sungas." The Sri Lankan chronicle, *Mahavamsa*, admits the existence of numerous monasteries in Bihar, Avadha, Malwa, and the surrounding areas during the reign of King Dutthaga-mani (ca. 101–77 BCE), which is synchronous with the later Sunga period.

Conclusion

It may not be possible to deny that Pushyamitra Sunga showed no favor to the Buddhists, but it cannot be said with certainty that he persecuted them. Though the Sunga kings, particularly Pushyamitra, may have been staunch adherents of ortho-dox Brahmanism, they do not appear to have been as intolerant as some Buddhist texts have shown them to be. The only thing that can be said with certainty on the basis of the stories told in Buddhist texts about Pushyamitra is that he might have withdrawn royal patronage from the Buddhist institutions. This change of circum-stances under his reign might have led to discontent among the Buddhists.

It seems that as a consequence of this shifting of patronage from Buddhism to Brahmanism, the Buddhists became politically active against Pushyamitra and sided with his enemies, the Indo-Greeks. As H. Bhattacharyya and col-leagues (1953) suggest, this might have incited him to put the Buddhists down with a heavy hand. Thus if in some parts of Pushyamitra Sunga's kingdom a few monasteries were at all pillaged, it must be seen as a political move rather than a religious one. Moreover, in such cases the complicity of the local gover-nors also cannot be ruled out.

Jayaswal (1923) has referred to another interesting aspect of the declaration made by Pushyamitra Sunga at Shakala, the capital and base of Menander. Accord-ing to him, the fact that such a fervid declaration was made not only at a place that

was far removed from the center of the Sunga regime but also in the capital city of his arch enemies, points to reasons motivated by political considerations.

After Ashoka's lavish sponsorship of Buddhism, it is quite possible that Buddhist institutions fell on somewhat harder times under the Sungas, but persecution is quite another matter. Thus it would be fair to say that where the Buddhists did not or could not ally themselves with the invading Indo-Greeks, Pushyamitra did not beleaguer them. In any case, after the end of the Sunga Dynasty, Buddhism found congenial environment under the Kushanas and the Shakas, and it may be reasonable to assume that Buddhism did not suffer any real setback during the Sunga reign even if one could see some neglect or selective persecution of Buddhists. As Lamotte writes: "Far more than the so-called persecution by Pushyamitra, the successes of the Vishnuite propaganda during the last two centuries of the ancient era led the Buddhists into danger, and this was all the more serious in that it was a long time before its threat was assessed" (1988: 392–393).

References and Further Reading

Aryamanjusrimulakalpa, The. K. P. Jayaswal, ed. and trans. *An Imperial History of India*. Lahore: Motilal Banarsidass, 1934.

Bhattacharyya, H., et al., eds. *Cultural Heritage of India*. 2nd rev. ed. Vol. II. Calcutta: Ramakrishna Mission Institute of Culture, 1953.

Chakravartty, Gargi. "BJP-RSS and Distortion of History." In *Selected Writings on Communalism*, edited by P. Lahiri, 166–67. New Delhi: People's Publishing House, 1994.

Devahuti, D. *Harsha: A Political Study*. 3rd rev. ed. Delhi: Oxford University Press, 1998.

Divyavadana, The. P. L. Vaidya, ed. Darbhanga: Mithila Institute of Post-Graduate Studies and Research in Sanskrit Learning, 1959.

Elst, Koenraad. "Why Pushyamitra Was More 'Secular' than Ashoka." 2005. http://koenraadelst.bharatvani.org/articles/ayodhya/pushyamitra.html (accessed June 1, 2010).

Ghosh, N. N. "Did Pusyamitra Sunga Persecute the Buddhists?" In *B.C. Law*, Vol. I, edited by D. R. Bhandar. Poona, India: Bhandarkar Research Institute, 1945.

Ghoshal, U. N. *A History of Indian Public Life*. Bombay: Oxford University Press, 1966.

Jayaswal, K. P. "An Inscription of the Sunga Dynasty." *Journal of the Bihar and Orissa Research Society, Patna* 10 (1923): 203.

Jayaswal, K. P. "Revised Notes on the Brahmin Empire." *Journal of the Bihar and Orissa Research Society, Patna* (September 1918): 257–65.

Lamotte, E. *History of Indian Buddhism: From the Origins to the Śaka Era*, trans. by Sara Webb-Boin. Louvain-la-Neuve, Belgium: Insitut Orientaliste: 1988.

Marshall, J. H. *A Guide to Sanchi*. 3rd ed. Delhi: Manager of Publications, 1955.

Marshall, J. H. *A Guide to Sanchi. Taxila*, Vol. I. Delhi: Bharatiya Publishing House, 1975.

Mishra, P. K. "Does Newly Excavated Buddhist Temple Provide a Missing Link?" *Archaeology, A Publication of the Archaeological Institute of America* (April 2001). Available at http://www.archaeology.org/online/news/deorkothar/index.html (accessed June 1, 2010).

Mitra, R. C. *The Decline of Buddhism in India, Santiniketan*. Birbhum, India: Visva-Bharati, 1954.

Mukhopadhyaya, S., ed. *The Aāokāvadāna*. New Delshi: Sahitya Akademi, 1963. Translated by John S. Strong as *The Legend of King Ashoka: A Study and Translation of the Aāokāvadāna*. Princeton, NJ: Princeton University Press, 1983.

Negi, J. S. *Groundwork of Ancient Indian History*. Allahabad, India: Allahabad University Press, 1958.

Raychaudhury, H. C. *Political History of Ancient India: From the Accession of Parikshit to the Extinction of the Gupta Dynasty*. Calcutta: University of Calcutta, 1923.

Sastri, Haraprasad. *Journal of the Asiatic Society of Bengal, Calcutta*, (1910): 259–62.

Sircar, D.C., ed. *Select Inscriptions Bearing on Indian History and Civilization*, Vol. I, 2nd rev. ed. Calcutta: University of Calcutta, 1965.

Tāranātha's History of Buddhism in India. Lama Chimpa and A. Chattopadhyaya, trans. Simla: Indian Institute of Advanced Study, 1970.

Thapar, R. *Ashoka and the Decline of the Mauryas*. New York: Oxford University Press, 1991.

6

The Shroud of Turin is actually the wrapping shroud of Jesus.

PRO   Justin Corfield
CON   Thaddeus Nelson

PRO

The Shroud of Turin, held at the Turin Cathedral in northern Italy, has long been thought to be the burial cloth of Jesus Christ. It was used to wrap around Jesus's body after the Crucifixion and was left in his tomb when he "rose from the dead." There are several plausible theories over how it has survived, and indeed its history dates back to the late Middle Ages and is known with a fair degree of certainty. There is also much scientific evidence for the probability that it was genuine, with the main evidence against it being the carbon dating conducted in 1988 of some pieces.

The Jewish tradition of burial was to wrap the body in a cloth and bury the person on the same day of death. In the case of Jesus it was after his body had been taken down from the cross and his body was placed in the tomb prepared for him. In the New Testament of the Bible, all four Gospels (Matt. 27:59; Mark 15:46; Luke 23:53; John 19:40) refer to the body of Jesus, after it was taken down from the cross, as being wrapped in a "linen cloth"—Matthew mentions in addition that the cloth was "clean," and John mentions "cloths" (plural). Following the Resurrection, Mark refers to the sight of a man dressed in a white robe (16:5), with Luke referring to two men in "dazzling cloths" (24:4). John adds an extra piece of information that when Simon Peter and another disciple entered the empty tomb, they saw "the linen wrappings lying there. And the napkin that was about his head, not lying with the linen clothes, but wrapped together in a place by itself" (20:6–7). There are no other references in the Bible to the burial shroud. However, given the nature of the death of Jesus and his "rising from the dead," it seems likely that one of his followers would have kept the cloth. There have been numerous other surviving relics connected with Jesus, but the Shroud of Turin has been the most studied of these.

In about 1355, Geoffrey (or Geoffroi) de Charny, Lord of Lirey, in France, seems to have first publicly showed a cloth that he claimed was the Shroud of Jesus. He was a French landowner and soldier and was involved in much of the fighting at the start of the Hundred Years' War. He defended Tournai against

103

the English in 1340, and two years later fought at the Battle of Morlaix, again fighting the English. Five years later, he took part in a Crusade, going to Smyrna (modern-day Izmir, Turkey), and in 1349 was involved in a French attack on Calais. Captured by the English, he was taken to England as a prisoner, but seems to have been ransomed soon afterward, returning to France and writing a book on chivalry, which gained attention in scholarly circles. Geoffrey de Charny was killed by the English at the Battle of Poitiers in September 1356, and it is known that his wife publicly displayed the shroud soon afterward.

Although many sources cite his wife as being the person who first displayed the shroud, the earliest medieval written mention of the shroud dated from 1389, in a memorandum of the Bishop d'Arcis of Troyes. This refers to the shroud having been exhibited some "thirty-four years or thereabouts" before-hand, putting it within the lifetime of Geoffrey. The relatively specific mention of "thirty-four years" certainly ties in with the last years of Geoffrey's life—the battle of Poitiers being one of the major events in French history at that time. Seeming to confirm the connection between Geoffrey de Charny and the shroud, in 1855 a small pilgrim's medallion from the period was found and taken to the Muse'ee de Cluny in Paris. It shows the arms of Geoffrey de Charny and his wife Jeanne de Vergy and has a small depiction of what could be the shroud. It seems to indicate that some people might well have gone to the shroud as pilgrims in order to pray there for inspiration and would then get a medallion as a souvenir. Although Geoffrey may have been the first to exhibit the shroud, it is easily possible that it might have belonged to his wife.

Geoffrey and his wife's son, also called Geoffrey de Charny, inherited the shroud, which he exhibited at Lirey in 1389—the event that led to the bishop's memorandum. His daughter Margaret de Charny gave the shroud to Louis I of Savoy in 1453, and it was owned by the House of Savoy (later Kings of Italy) until the 20th century. The great-grandson of Louis, Emmanuel Philibert, moved the shroud to Turin in 1578, and his son Charles Emmanuel I, started plans for a spe-cial chapel for the shroud, which was installed in the Guarino Guarini Chapel in 1694. It was then exhibited at the marriages of various members of the House of Savoy—in 1737, 1750, and 1775—and in 1821 it was displayed to mark the acces-sion of Charles Felix as King of Savoy, and in 1842 to mark the marriage of Victor Emmanuel II, who in 1861 became the first king of Italy. After negotiations which started in 1973, in 1983 it was finally bequeathed by the great-grandson of Victor Emmanuel II, King Umberto II of Italy (in exile since 1946) to Pope John Paul II, and although it has remained in Turin Cathedral, it is the property of the papacy.

If the history of the shroud is known for certain back to 1355, the historical mystery surrounds what happened to it during the 13 centuries after the Crucifix-ion, if it is genuine. John's Gospel clearly describes the shroud, so there is the probability that it might have been taken by one of the followers of Jesus. If so, it is likely that it would have been treasured by the early Christians, but this is
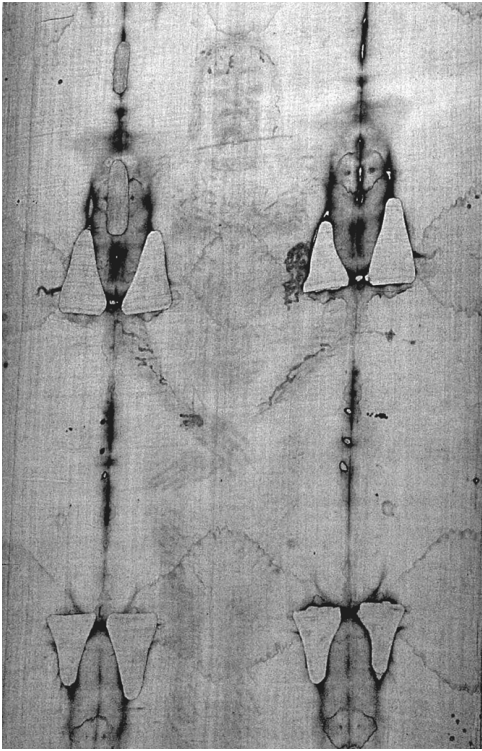
The Holy Shroud, a fourteen foot-long linen revered by some as the burial cloth of Jesus, on display at the Cathedral of Turin, Italy. (AP Photo/Antonio Calanni)

all supposition. Most historians now identify the shroud with the Byzantine relic known as the Mandylion or "Image of Edessa," held in Constantinople until the city was sacked in 1204. It was a cloth that purported to have on it the face of Jesus Christ, and there is some evidence—written and visual—that the face on the Mandylion and the shroud were similar.

The Mandylion had been taken to the city of Constantinople in September 944 by pilgrims from the city of Edessa (modern-day Urfa), in southern Turkey, where it was claimed to have been held since the sixth century. The city has some links with the Bible, with some historians identifying it as the biblical city of Ur, although this is disputed by the vast majority of historians. Located along a major caravan route, in common with many other places in the region, Edessa had a small Christian community. A Roman headquarters, it had been used by the Emperor Valerian in 260, and in the fourth century, Saint Ephram had lived there, founding a school of theology. The people in Edessa followed the Nestorian beliefs, which were found to be at variance, theologically, from those of the Byzantine rulers, and the school was closed in 439. The city was captured by the Sassanids in 605, retaken by Heraclius, and then captured by the Arabs in 639 who brought Islam to the city. What happened to the shroud during this period is a matter of pure conjecture, but its transfer to Constantinople in 944 is mentioned in contemporary Byzantine documents.

The contemporary description of the Mandylion was that it was a picture of Christ "which was not made by human hands." There is also a reference made by the rusader Robert de Clari who described a "figure" on a cloth or shroud, which was held at the Church of Saint Mary at Blachernae in Constantinople and was shown to the public each Friday. The original account by Robert de Clari still survives in the Copenhagen Royal Museum. However, with the destruction of Constantinople during the Fourth Crusade in 1204, the relic—like many others—was lost.

To link the shroud with the Mandylion rests not just on the few necessarily vague written accounts. The face on the shroud is well known, and it shows the image of a man with a fairly distinctive face and beard, with several easily noticeable facial features. There are some images of the Mandylion surviving, and all these show a similar face—indeed they only show the face. This has been easily explained by the shroud having been folded in half, and then in half again, with the face in the center of the exposed piece of cloth. It would then have been mounted on a board and a lattice covering put on top to draw attention to the face, as seems to be the case in the images of the Mandylion. The folding of it seems likely, as the shroud shows the naked image of a man, and the Byzantines would probably have hidden this, especially with their belief that it was of Jesus. The earliest surviving image of the Mandylion dates to about 1100 CE and is in a fresco above an arch in the Sakli "Hidden" Church in Goreme, in Cappadocia, in central Turkey. There are also two images of the Mandylion from the 12th century. One is in a Serbian church at Gradac in modern-day Croatia, and the other is at Spas Neriditsa near Novgorod in the Russian Federation, with a later one surviving at a monastery at Studenica, Serbia, dating to the 13th century. A Byzantine coin from 945 shows an image of Christ similar to that on the shroud, and there is also a 12th-century mosaic portrait of Jesus in the cathedral at Cefalu, in central Sicily. However, it should also be pointed out that a coin from the reign of Justinian II (reigned 685–695 and again 705–711) has a similar portrait of Jesus, as does another Byzantine coin of 692. This seems to imply that the image of Jesus resembling the face on the shroud was well known long before the Mandylion came to Constantinople.

If the Mandylion and the shroud are the same, there are a number of possible connections that tie Geoffrey de Charny's family to the sacking of Constantinople in 1204. One tradition ties it to Hugh de Lille de Charpigny, who was present at the sacking, and it later ended up with lands at Aegion in Greece, somehow passing into the de Charny family, probably through his friend and companion-in-arms Guillaume de Champlitte, who was also at the sacking of Constantinople and whose wife, Elizabeth de Mont St. Jean, was the sister of Pons de Mont St. Jean, the great-grandfather of Geoffrey de Charny. Another possibility was that it might have come into Geoffrey de Charny's family through his wife Jeanne de Vergy. One of her great-great-great grandfathers was Otho de la Roche, who also took part in the sacking of Constantinople in 1204. A third link exists, and this involves the Knights Templar.

The Order of the Poor Fellow-Soldiers of Christ and the Temple of Solomon, better known as the Knights Templar, was a secretive Crusader military order established in 1119, with strong ties to central France, from where many of its leaders came. The order was known to possess many secrets and had been established in Jerusalem with a particular interest in holy sites, having carried out its own excavations at the site of the temple in Jerusalem. There has long

been much controversy about a "head" known as "Baphomet," which was used in some of their meetings and revered by the knights. When the order was destroyed in 1307, interrogations of the knights were unable to reveal exactly what "Baphomet" was—certainly no trace of it has been found, and historians and pseudo-historians have debated whether or not it could refer to the shroud, or even the head of Jesus or John the Baptist. There have been many theories, and one curious one worth noting is that by authors Lynne Picknett and Clive Prince (1994). In their book *Turin Shroud—In Whose Image?*, Picknett and Prince have gone as far as to suggest that it was the image of the head of Jesus used by Leonardo da Vinci in the making of the Shroud of Turin—using early photography—meaning that the shroud is a late medieval creation, but it does actually show the head of Jesus.

There is one crucial piece of evidence for case of the Templars having the shroud, or at least having access to it. This comes from a Templar preceptory in Templecombe, Somerset, England, where Molly Drew, during World War II, discovered an old panel painting that was revealed after the plaster had fallen. It showed the image of a Christ figure, with a face similar to that on the shroud. Given that the order was suppressed in 1307, and with carbon dating placing Drew's panel back to about 1280, the link between the shroud and the Templars seems possible. Holger Kersten and Elmar Gruber (1994), in their book on the shroud, suggest that the wooden piece could have been part of the box in which the shroud might have been kept.

The Templars were involved in the sack of Constantinople, and the connection between Geoffrey de Charny and the Knights Templar is close, although he himself was not a member of the order. When the last grandmaster of the Templars, Jacques de Molay, was burned at the stake outside Notre Dame Cathedral, Paris, on March 18, 1314, his close aide who was burned with him was Geoffrey de Charnay (*sic*), the preceptor of Normandy for the Knights Templar. Historian Noel Currer-Briggs has traced that the man burned with Jacques de Molay was the uncle of Geoffrey de Charny, who put the shroud on public display in about 1355.

Whether the shroud was held by the de Charny family since the sacking of Constantinople, or whether it was held by his wife since that time, or even whether it was owned by the Knights Templar, it is clearly possible that Mandylion and the shroud could be the same item; and given their similarities, it seems probable that they were the same. This, therefore, manages to push back the existence of the Shroud of Turin to as far back as the sixth century, when it was celebrated as being in Edessa. The written records cannot provide any more information, but there is plenty of evidence on the shroud itself that provides far more positive evidence that it was the burial cloth of Jesus Christ.

The image on the shroud has been minutely examined and found to clearly not be made by paint. There are several theories over how it could have been

formed. The microanalyst Walter McCrone (1990), from Chicago, has maintained that the image on the shroud could have been made from iron oxide pigments using gelatin as a binding medium. The major problem with this is that the image on the shroud represents an extremely tall man, and some of the features show an odd perspective. The most prevalent theory about these questions is that the image on the shroud is capable of being formed naturally when a body is covered in various substances—as would have happened to the body of Jesus when it was placed in the tomb (John 19:40). Historian and author Holger Kersten, in his work with Elmar Gruber, ground aloe and myrrh, and the resulting experiment showed that it is possible to imprint the image of somebody onto a shroud-like garment. Kersten carried out several experiments, and this tended to back up his theory and also show how the image shown in the shroud was slightly misshaped, possibly resulting from the cloth closely following the contours of the body—the body being lain on the cloth, which was then used to cover it up. The cloth would then be stitched to help people carry the body of the deceased. The fact that the cloth follows the shape of the body explains the reason why the figure on the shroud was so tall, the image not being a two-dimensional image. Given that the exact method of treatment of the body of Jesus, and indeed others in Jerusalem during the same period, is not known for certain, it is clearly possible for the image to come from the person buried in the shroud without any supernatural significance being given to the existence of this image.

As the nature of the image can be easily explained, what it depicts needs close examination. There are many parts of the image that tally with the Gospel descriptions of the Crucifixion of Jesus, but there are discrepancies with medieval beliefs that are important, as some feel the shroud was created in the late Middle Ages. The most obvious is that the nails used in the crucifixion of the man whose image is left on the shroud were nailed through his wrists. Although church paintings of the period, and indeed for many centuries afterward, show the nails going through the palms of his hands, this would have been impossible in an actual crucifixion, as the palms were not capable of taking the weight of someone's body. Parts of the bones of a crucifixion victim were found in June 1968 to the north of Jerusalem in a burial ground, which can be dated to the time of Jesus. On these it is quite easy to see that the nails passed through the wrists. The next controversial point about the hands was that the thumbs of both hands couldn't be seen. This is because when the nail passed through the wrist, it led to a contraction of the thumb, evidence of which has been noted when experiments were done on amputated limbs. If the figure on the shroud had been faked, it seems unlikely that the forger would have both been able to transpose the nailing from the palms to the wrists and be aware of the effects of this on the thumbs. This therefore suggests that the image is of someone who was either crucified by the Romans or in a manner similar to that used by the Romans. However, there is a problem of rigor mortis. If rigor mortis had set in, it would,

obviously, have been impossible to rest the hands of the figure in the shroud over his groin. This again tallies with the biblical account that the body of Jesus was taken down very soon after he had died. It has also led to theories that Jesus was thought to be dead when he was brought down from the cross—either having fainted or having been drugged.

A number of writers have been able to observe many other pieces of evidence of the image. There is clear evidence of flagellation, with the image on the shroud clearly having been naked at the time. It has also been possible to spot evidence of other parts of the biblical account of the death of Jesus—the Crown of Thorns and the spear in the side. These all tally with the Gospel accounts.

Historians and scholars have also studied the weave of the cloth. The method of weaving in a herringbone pattern was common in Syria at the time of Jesus, but unknown outside that region. Although this does suggest that the cloth came from Syria, it still does not prove that it was that of Jesus, although once again it points away from being made in medieval France. However, the pattern of the weave as well as the method of lying the body of Christ on the cloth is shown in a fresco in a church in Nerezi, near Skopje, the capital of Macedonia (the former Yugoslav Republic of Macedonia), dating from 1164, and also from the Codex Pray, a prayer book compiled in Budapest in 1192; as well as in an *epitaphion* from Thessaloniki in Greece, dating from the 14th century, all pointing to the fact that details of a cloth similar to the Shroud of Turin were well known for centuries. This once again suggests the link between the Mandylion and the shroud.

Historian Ian Wilson has been keen to prove the validity of the shroud or whether it was made after the time of Jesus. His book, *The Turin Shroud*, first published in 1978 and then enlarged as *The Blood and the Shroud: New Evidence That the World's Most Sacred Relic is Real* (1998), provides much of the detail on the shroud and the possible reasons for it being genuine. For most skeptics, the main reason for doubting its genuineness rests on a series of carbon dating tests carried out in 1988. These have been seen by many as proof that the shroud dates to the late Middle Ages. This would therefore mean that the cloth held by Geoffrey de Charny, first shown by either himself or his widow, is not the same one as currently in Turin Cathedral, suggesting that at some stage the medieval "original" (which may, or may not have been that of Jesus) was replaced by a late medieval "copy."

There is a strong belief in the infallibility of science, and many commentators have seen the carbon dating as proof that the shroud could not be that of Jesus. However, many serious doubts have arisen as to the accuracy of the carbon dating. The first reason for querying the carbon dating concerns the part of the shroud that was removed to be tested. As the Turin authorities were loath to let any significant part of the shroud be burned for carbon testing, the small piece that was tested—and destroyed during the testing—was from the edge of the cloth. This immediately raises the query whether it might have been a part of a

The Shroud of Turin Research Project

In early 1970s, a group of scientists, mostly from the United States, formed the Shroud of Turin Research Project (STURP). In 1978, 24 scientists from STURP spent five days gathering evidence from the shroud, resulting in a 1981 report that put to rest many of the theories of spurious origins:

No pigments, paints, dyes or stains have been found on the fibrils. X-ray, fluorescence and microchemistry on the fibrils preclude the possibility of paint being used as a method for creating the image. Ultra Violet and infrared evaluation confirm these studies. . . .

Microchemical evaluation has indicated no evidence of any spices, oils, or any biochemicals known to be produced by the body in life or in death. . . .

We can conclude for now that the Shroud image is that of a real human form of a scourged, crucified man. It is not the product of an artist.

Source: Shroud of Turin Web site. "A Summary of STURP's Conclusions." Available at www.shroud.com/78conclu.htm. (accessed June 1, 2010)

Renaissance repair—the shroud having been repaired on a number of occasions. However, the real problem over the carbon dating was the lack of a "control" test whereby material of a known date was burned and the results compared with those from the shroud.

The first pieces of the shroud removed for examination on November 24, 1973 were studied in detail by the Belgian textile expert Gilbert Raes, director of the Laboratory for Textile Technology in Ghent. At that time the amount of the shroud that would have to be removed for carbon dating was too big for it to be considered. However, 15 years later, technology had advanced such that one small piece was removed on April 21, 1988. This piece was cut away by Giovanni Riggi, a specialist in microscopy, but the method of testing it had changed. Initially, parts were going to be sent to seven laboratories. Subsequently, it was changed and a new protocol was reached with the British Museum, London, acting in a coordinating role. The tests involved pieces from the shroud being tested against two control specimens from cloth of a known age. As a result, as soon as the piece was detached from the shroud, in the full blaze of publicity and in front of witnesses, it was then taken into an adjoining sacristy where it was cut into three pieces, and the control cloth was placed in nine small tubes to be sent to three radiocarbon laboratories. The first three—an actual piece, along with two "control" pieces—were tested at Tucson, Arizona, with the next three tested in Zurich, and then the last three pieces were tested at the radiocarbon laboratory in Oxford, England. The various bodies agreed to communicate the results to the Vatican ahead of publicizing the details, and on October 13, 1988, at a press

conference held in London, Edward Hall from Oxford, his chief technician Robert Hedges, also from Oxford, and Michael Tite from the British Museum in London announced the results. All they had with them was a blackboard upon which the dates "1260–1390" were written. Tite explained to the press that the radiocarbon dating had come up with that period of years to a 95 percent degree of probability, with the shroud's raw flax being made into linen possibly in or around 1325. Although many people feel that the carbon dating has proved that the shroud, in spite of all the circumstantial evidence tying it to the Mandylion, was made in the late medieval period, the scientific account of the carbon dating was not published until February 16, 1989, in the scientific journal *Nature* (Damon 1989).

There were, however, several problems involving the carbon dating. The first was raised by the right-wing Brother Bruno Bonnet-Eymard of the Catholic Counter-Reformation in the Twentieth Century. He pointed out that while the piece of the shroud was removed in front of cameras, the putting of sections of the piece into canisters for testing was done in secret by Tite and an elderly cardinal, and he accused Tite of having "switched" the samples. Kersten and Gruber suggest that this was because the image on the shroud was not "super-natural." They felt that the blood on the shroud proved that Jesus did not die on the cross but in the grave—or at any rate on the shroud. Others have suggested that the washing of the body of Jesus, necessary for the image to form on the shroud, might have been the cause of the blood.

However, Kersten and Grueber had a few other reservations over the whole carbon-dating process. As well as the possible switching of the pieces of the shroud and the two "control" cloths, they had queries over exactly which cloth was used as the control. The dates given by the three radiocarbon laboratories varied considerably, not just with the cloth from the shroud—presuming it had not been swapped—but with the tests on the cloth from the "controls." One of the pieces used in the control was stated, in *Nature* magazine, as coming from the cope of Saint Louis d'Anjou, great nephew of King (later Saint) Louis IX. However, despite efforts by Kersten and Grueber to track down the cloak, which had last been restored in 1965, they were not able to discover where it was now located. Kersten and Grueber went further and suggested that rather than testing the shroud against medieval cloth, it would have been important to test it against ancient cloth and cloth of a known date. They themselves had found plenty of ancient cloth from the Middle East at the Victoria and Albert Museum in London that could have been used.

Ian Wilson's (1998) criticism of the carbon dating is different. He pointed out that the carbon dating left many questions unanswered. The weave of the cloth of the shroud was different from the controls, and as each cloth was photographed by the various laboratories—many without scales so it is impossible to determine, to any degree of accuracy, the size of the pieces being tested—it

would have been possible for even lay observers to guess which parts belonged to the shroud and which did not.

This leads to the study of pollen found on the Shroud of Turin. Trying to locate pollen was the idea of Max Frei, who was head of the Zurich police laboratory. He had written on the flora of Sicily and used clean strips of adhesive tape to remove pollen from parts of the shroud. By March 1976 he had been able to differentiate between pollen from 49 different plants on the shroud—one being from the cedar trees of Lebanon, and others from halophytes, plants that need a very high salt content that would have flourished around places such as the Dead Sea. He was able to prove that the shroud does have pollen of plants that are found only in the Middle East, and many found largely in the Holy Land. This proved that the cloth had been, for part of its history, in the Holy Land—presumably prior to it being taken to Edessa. The accuracy of his work was proven by the location of a rice pollen. This was easily explained because the Shroud of Turin was displayed from the balcony of the castle of Vercelli in northern Italy in 1494 and 1560, which at that time was in the center of the main rice-growing region in Europe.

Undoubtedly, questions remain about the shroud, and there are anomalies and gaps in the story of how it moved from Jerusalem to Edessa, how it survived the sacking of Byzantium, and how it ended up in France. However, all the theories of it being fake—as a painting or an early photographic image—can either be totally disproven or leave far more questions unanswered. For most people, including the hundreds of thousands of pilgrims who have flocked to Turin each time the shroud has been placed on public view, the shroud remains the burial cloth of Jesus.

References and Further Reading

Craik, Jamie E. "The Templars, the Shroud, the Veil and the Mandylion." *Templar History Magazine* 1, 1 (Fall 2001): 17–20.

Currer-Briggs, Noel. *The Shroud and the Grail*. London: Weidenfeld and Nicolson, 1987.

Goldblatt, Jerome S. "The Amazing Facts of the Turin Shroud." *Bulletin* (April 5, 1983).

Gove, Harry E. *Relic, Icon or Hoax? Carbon Dating the Turin Shroud*. Bristol, UK: Institute of Physics Publishing: Bristol, 1996.

Kersten, Holger, and Elmar R. Gruber. *The Jesus Conspiracy: The Turin Shroud and the Truth about the Resurrection*. Shaftesbury, UK: Element, 1994.

Laidler, Keith. *The Divine Deception: The Church, the Shroud and the Creation of a Holy Fraud*. London: Headline, 2000.

McCrone, Walter C. "The Shroud of Turin: Blood or Artist's Pigment?" *Accounts of Chemical Research*, 23:3 (1990) 77–83.

Picknett, Lynn, and Clive Prince. *Turin Shroud—In Whose Image?* London: Bloomsbury Books, 1994.

Damon, P. E., et al. "Radiocarbon Dating of the Shroud of Turin," *Nature* 337:6208 (February 16, 1989), 611–615.

Walsh, John. *The Shroud*. London: W. H. Allen, 1964.

Wilson, Ian. *The Blood and the Shroud*. London: Weidenfeld and Nicolson, 1998.

Wilson, Ian. *The Turin Shroud*. Harmondsworth: Penguin, 1979.

Wilson, Ian, and Barrie Schwortz. *The Turin Shroud: The Illustrated Evidence*. London: Michael O'Mara Books, 2000.

CON

The Shroud of Turin has become a highly recognizable symbolic object, tied to Jesus Christ, his death, and Resurrection as told in the Christian Gospels. The shroud is a piece of intricately woven linen cloth, roughly 4.3 by 1.1 meters in size. When unfolded, it appears to contain an image of a crucified body. Supposedly, the shroud was used to wrap Jesus after his Crucifixion and remained behind after his Resurrection. This story holds a place in the hearts of many believers, offering them a direct tie to important events lost nearly 2,000 years ago. While many are willing to accept this on faith, others have exposed the shroud to scientific inquest, meant to understand the actuality of such claims. In almost every case, the shroud has been shown to be either a forgery or an artistic piece passed off as real. In response to these results, many who believe that it is the burial cloth of Jesus have presented arguments to attempt to disprove the scientific studies. These arguments are often not valid, and no arguments have actually been presented as positive evidence that the shroud was tied to Jesus.

Evidence against Authenticity

Perhaps the first and most important piece of evidence against the authenticity of claims that the Shroud of Turin was used to wrap Jesus Christ in death is the history of the object. Clear points of concern surface from the early history of the shroud that illustrate it was recognized as a fraud from its first showing. The record of the shroud begins in the town of Livey, France, in 1355 CE, where it was owned by Geoffrey de Charny. He offered no explanation for the origin of the shroud but simply claimed that it was the burial cloth of Jesus. Near the end of 1389, Bishop Pierre d'Arcis wrote to Pope Clement VII, telling him of an investigation launched by his predecessor, Bishop Henri de Poitiers, and the final results of his study of the shroud. Bishop de Poitiers had told Bishop d'Arcis that an artist had confessed to him that he had created the work and that

it was not the actual shroud in which Jesus was buried. Pope Clement VII, after weighing the evidence, declared that the shroud was a fake and that it could be displayed as a representation, but not as an authentic relic.

It would seem that after such a confession, the tale of the shroud should have ended in the late 14th century. However, Charny's granddaughter Margaret, resumed showing the shroud as authentic during the early and middle 15th century. During this period, there were additional questions of the shroud's authenticity, showing that it was not held as a reputable item during Margaret's ownership either. In fact, her husband recorded the shroud in his papers only as an image, not as the actual burial cloth of Jesus. In 1453 Margaret willingly sold the shroud to the Royal House of Savoy. While in their care, it was threatened by a fire in 1532, an event that will be important in further analysis of the shroud's authenticity. When the Royal House moved its capital to Turin in 1578, the shroud went with it. From this location, where it remains today, it gained its familiar name. In 1983 the Savoy family gave the shroud to the pope and the Catholic Church, which has not stated that it is authentic and instead leaves such matters to the faith of individuals.

This brief history of the Shroud of Turin offers a number of important points concerning the issue of authenticity. When Pope Clement VII considered the shroud, one point that his decision was influenced by was the fact that Geoffrey de Charny could not provide an explanation for its origins. This was a very forward-thinking methodology. To a modern archaeologist, a record of an artifact's origin and successive owners is its provenance. Modern scholars rely heavily on provenance, as did Pope Clement VII. Without such records, it is impossible for scholars to tie objects to their origins. What this means for the shroud is that not only is it impossible to state with a high degree of accuracy that it is authentic, but also there is no way to state that it even existed before 1355. Within the light of a confessed forger, this problem alone should cast doubt on the proposed linking of the shroud to Jesus around 30 CE.

Factors for Proving Authenticity

There remain many who still claim the shroud is in fact the burial cloth of Jesus. To best address these continued claims, we can easily look at what evidence would support such a position and see if it exists. Unfortunately, DNA from Jesus can't simply be found and compared to the portions of the shroud that appear to show blood. Since there is no such method of directly linking the shroud and Jesus, its epistemology must be approached. This means looking at how the shroud fits into the already established body of scientific knowledge about the past, based on the evidence that can be gathered.

One way the shroud would have to fit within the existing knowledge of Jesus is that it should date in creation to the early first century CE. This is so that

its creation might predate the Crucifixion. Until 1989 this was an issue of great contention because it was very difficult to determine exactly when it was made. By the 1980s, the possibility of carrying out carbon-14 dating on the shroud became a reality, with more efficient methods that would not require the destruction of a large portion of the cloth. Carbon-14, or radiocarbon, dating measures the predictable decay of a specific type of radioactive carbon atom found in all organic material, carbon-14, over time to find an accurate date for the death of living things. The results, published in 1989, showed that the plants used in creating the fibers composing the shroud were most likely harvested between 1260 and 1390 CE. Such a date is strong evidence against the shroud being used to wrap Jesus's body.

Believers of the shroud's authenticity have not let the results of carbon-14 dating keep them from supporting a much older age. There have been numerous arguments as to the validity of these findings based on potential inaccuracies in the radiometric dating. The best-known claim of problematic carbon-14 dating focuses on the fire the shroud survived in 1532. According to proponents of the shroud's authenticity, the heat and chemical exposure in the fire would have contaminated the shroud, providing a more recent date through carbon-14 testing. Such a hypothesis is problematic for two reasons. First, carbon-14 dating of burnt or charred materials is a common practice in archaeology and can produce accurate results. Next, H. E. Gove (1996) makes the point that even if the fire were to somehow introduce additional carbon-14 to the shroud, in order for carbon introduced in 1532 to skew tests in the 1980s by 1,100 to 1,200 years, 86 percent of the carbon in the shroud would have had to originate in the 1532 fire. Such an inclusion of carbon is not only incongruous with scientific understanding of fires, but it also would leave an obvious addition of material that has never been found on the shroud. There is thus no reason to believe that the shroud changed its carbon makeup in 1532 in a way that would meaningfully alter the carbon-dating results of the 1980s.

While the 1532 fire could not have altered the shroud's dating, there are other hypotheses put forward in defense of a first-century dating, other possible influences on the outcome of the tests. One popular explanation is that the samples of the shroud taken for dating included threads from a modern repair and were thus skewed toward later dates by the inclusion of younger materials. This theory does not fit well with the actual events involved in sample selection. The process involved the oversight of two textile experts and inspection of the sample under magnification. Even the most modern stitching at the time of the tests would have been visible under a microscope and would certainly have been noticed by the experts. There is no actual evidence to support the existence of such thread in the area tested either, simply the guess of those who favor an earlier date for the shroud's manufacture. Without such evidence, researchers must rely on the opinion of the textile experts and not the guesswork of shroud enthusiasts.

An additional source of contamination, and a subsequent misleading carbon-14 analysis, that has been proposed is microscopic organic compounds, such as bacteria, on the surface of the shroud. Such a defense is erroneous based on both the methods used in analysis and in the ability of such materials to significantly alter the dating of the shroud. The carbon-14 testing was carried out at three separate universities, in Arizona, England, and Switzerland. Each facility used methods of cleaning the cloth to remove foreign contamination. Between all three sites, any significant source of contamination would have been removed by one method of cleaning or another. As with the 1532 fire, a large proportion of the sample, 64 percent, would have to be composed of modern contamination to turn a first-century date into a date in the 13th or 14th century. This would leave only 36 percent of the sample as the actual shroud fabric. The cleaning would not have left anywhere near this level of contamination. Even if they had, it would not have been capable of removing all of it, and certainly such a large amount of contamination would have easily been detected.

Claims that the radiocarbon dating of the shroud was inaccurate are a problematic example of ad hoc hypotheses, meant to explain away evidence contrary to one's desired outcome. There is no evidence to support any inaccuracies, and the methodology involved supports the position of the universities concerning the shroud's date. Methodologically, the analysis is robust. Three universities, using different preparation methods, arrived at similar dates, close enough to be beyond the likelihood of chance. To further assess the accuracy of the tests, each university tested three additional samples of cloth that were of known age. The results were similar to the shroud in the closeness among the universities' findings, but it is also possible from these data to say that the dates provided were accurate based on known information. This illustrates that the methods used were accurate and that the preparation and analysis of the shroud should be trusted as similarly accurate.

Basis of Arguments for Authenticity

With the work of Max Frei, supporters of the shroud's authenticity found some support for their position in pollen fossils. Supposedly these miniscule particles were lifted from the shroud's surface with tape and, when analyzed, they showed an origin in the Middle East. This would seemingly match the epistemological model of Jesus's life and place the shroud within the area of his burial. However, such agreement is not positive proof of a link between the two. It would offer no evidence of time frame or the owners of the cloth. This piece of the puzzle was also problematic in that it was not able to be repeated. In 1978 the Shroud of Turin Research Project (STURP) took another course of taped samples from the surface of the cloth. These lacked corroboration for the pollen fossils observed by Frei, calling into question his results as potentially missampled. It is important

to stress that even if such pollen remains were on the shroud, this at best can be seen to place it at some point in the region that Jesus came from, but not provide ample direct evidence of use of the cloth in his burial.

While the 1978 STURP tape samples have been used in studying potential pollen samples, this was not the original purpose in collecting them. STURP lifted 36 samples from portions of the cloth both within the image and in areas with no apparent image in an attempt to find support for the idea that the shroud was not formed through artistic means. The findings were that the samples did not provide evidence of man-made pigments or painting and that the image was created through some other means. This theory has been prevalent since at least 1898, when the shroud was first photographed. In these, it appeared that the shroud was a negative image, dark where it should have been light and light where it should have been dark. Proponents of the theory that the shroud was used to bury Jesus state that no artist would have been able to produce such an image. A final point to support the shroud's authenticity concerns details that appear in the shroud that artists in medieval Europe would not have known about. If any aspect of the shroud's formation is truly beyond the ability or knowledge of artists in the 13th and 14th centuries, this would raise the question of how exactly the shroud was made. The question that must be asked is if any of these claims is true.

Arguments that Medieval Artists Could Not Have Crafted the Shroud

Arguments that medieval artists could not have crafted the shroud generally fall into two categories: unknown historical points and anatomical accuracies. The most widely stated claim is that the points in the shroud where the crucified victim would have had nails placed are in his wrist, not the hand, as was prevalently illustrated in artistic representations. Modern scientific study has shown that placing the nail in the hand would not have worked, as they cannot bear the weight of an adult human. However, the image on the shroud only has one such nail, and it appears to actually be in the lower portion of the hand, not the wrist. Other supporters have claimed that the flagellations shown in the shroud would have been unknown in medieval Europe. This ignores contemporary artistic representations of just such marks. No aspect of the shroud's imagery appears to be outside of the knowledge of potential medieval artists, and while this does not prove that it is constructed by such individuals, it means that this cannot be precluded either.

A second line of reasoning used to counter claims that the shroud is only an artistic interpretation concerns the anatomical details preserved in the image of the figure. If anything though, the anatomy present indicates in various ways that the shroud is just such a human endeavor. Perhaps the most important point is that the proportions and layout of the body do not match those of real humans. The body is extremely disproportionate to actual anatomy, appearing

---

The Face on the Shroud: Jesus or Leonardo?

Recent scholars, looking at the face on the Shroud of Turin, have come to the conclusion that what created the image on the shroud had to have been a photographic process, as the shroud contains no pigments and the image is in the negative. However, this has led to a question about who would have had both the knowledge and the materials to create such an image, and, most importantly, just whose face was used. Lillian Schwartz of the School of Visual Arts in New York has scanned the face on the shroud and come to the conclusion that it has the same dimensions as the face of Leonardo da Vinci, one of the few people at the time of the shroud's supposed creation who would have had the knowledge of both human anatomy and the photographic process needed to create such a forgery, as well as the access to the materials to carry it off. Using an early photographic device, a *camera obscura*, and a sculpture of his face, da Vinci could have used silver sulphate to make the fabric sensitive to light, leaving a permanent, negative image of the face on the sculpture—possibly his own.

---

tall and lanky. This has lead to the proposition that Jesus, if it was his burial cloth, would have suffered from a genetic disorder called Marfan's syndrome (Nickell 1993).

A better explanation may be found in gothic art, which frequently depicted humans in such a drawn out manner. Additionally, it has been noted that the hair seems to hang down as if the figure were standing, not laying recumbent, as would have been the practice in first-century Jewish tombs. This observation does not require that the shroud would have been created artistically, but it does not speak well of the defense that the shroud is perfectly correct in physical details. Of similar concern is the position of the legs, one of which is shown straight and out of place with the associated footprint. If a real body had directly created the image, one would expect a flexed leg to match with the foot. A final point is the clarity of the image. It appears that the cloth was wrapped around a still body and then not moved. Besides the difficulty in placing the body without significant movement, after death the body would have settled and moved slightly as rigor

Enhanced photograph of what is believed by some to be the face of Jesus Christ as it was impressed in the Shroud of Turin. (Chiesa Cattolica Italiana)

mortis released. What the shroud actually shows is an idealized individual whose positioning and physical body fit better with an artistic interpretation than an actual human corpse.

If the details of the shroud do not preclude artistic work and the biology of the individual suggests it, STURP's findings that there were no pigments involved in the creation of the shroud should be questioned. This is exactly what Walter C. McCrone did in 1989, using better equipment and a lifetime of experience as a microanalyist (McCrone 1990: 77–83). Using STURP's sample tapes, McCrone was able to detect two common pigments, red ochre and vermillion, as well as other evidence of painting on the shroud. McCrone was able to show clear evidence of the use of pigments in both the body and bloody images. He was also able to determine that the variety of red ochre used was not available until after 800 CE, offering additional evidence for the earliest possible dating of the shroud's creation. The presence of pigments is not at all surprising, as the bloody parts of the image remain red to this day, something that would not be expected from actual blood, which turns brown over time. When assessed in light of the details present in the shroud, McCrone's analysis indicates an artistic origin.

Persistence of Pro-Authentic Argument

At this point, it is clear that the Shroud of Turin was likely created well after the time of Christ's burial and that it shows evidence of being a man-made image. Even in light of these facts, individuals still claim that the shroud must have been the burial cloth of Jesus. To bolster their position, they present the interpretation of the shroud as a negative image, popularized by photographs taken in 1898, and to the fact that the image does not penetrate through the fabric of the shroud as evidence that no one in the 14th century could create it. Part of the interpretation of this position lies in theories about how the shroud's image may have formed through contact with Jesus's body. Like all claims, these should be analyzed before they are blindly accepted.

Proponents of the body contact theory hold that direct contact with the herbs and oils or their vapors used in cleaning Jesus would have marked the fabric with an image of Jesus's body. This model is quite problematic, as vapors do not travel in a straight line, nor are they focused in the way that would be necessary to form the image of the shroud. Attempts to duplicate this method of formation have met with failure, creating blurry images completely unlike that of the shroud.

Supporters of the shroud's authenticity claim it could not have been created through painting. Scientists who have studied the shroud have presented evidence that shows this likely is not true. Beyond his findings of paint on the shroud's surface, McCrone was able to locate an instructional book from the 1800s explaining a 14th-century technique for creating images that may have been used to make the shroud. In fact, this method appears to have been meant

specifically for creating almost invisible images on cloth. This fits well with the date of forgery presented to Pope Clement VII. Joe Nickell (1983) added support to the theory that the shroud was painted by performing a trial of his own. By placing cloth over a model of a human and dabbing it with dye, he was able to present an image that matched the shroud. Nickell found that the images of both his reproduction and the actual shroud were not the true negative that shroud enthusiast claimed would be the result. The beard and hair of the individual actually appeared as positive images.

There is an epistemological argument against the idea that the shroud could not be a painting. This deals with how it fits into existing knowledge of Christian artistic traditions. As noted before, the body of Christ on the shroud fits better with gothic artistic tradition than actual human anatomy. The artistic representation of Jesus fits well with a long chain of development that began in the middle of the second century, when the first images of Christ appeared. Before this point, the anionic nature of early tradition prevented the creation of any divine images. A further step toward the shroud appeared in the sixth century, with the development of a theology of unmade images. These were icons supposedly created not by people, but by a divine act, a theory that will be explored more in depth at a later point. The shroud itself can be seen as a part of these artistic and theological traditions, which are also composed of artistic representations. This is not absolute proof that the shroud was a human-created item, but in conjunction with the evidence of paint, the impossibility of the vapor method of creation, and the proven existence of artistic methods for creating similar images, it is safe to say that the shroud was an artistic endeavor.

Basing on Belief

There remains one final argument that shroud supporters frequently fall back on when the overwhelming scientific and historical evidence is presented. Shroud proponents state that it was a unique item created through a onetime miraculous action that has never occurred since. In general, claims of divine origin and power are not considered scientific or acceptable as evidence in science, history, or archaeology. This is because they cannot be tested and disproved, a cornerstone of the scientific method as it is used across these disciplines. Such arguments are within the realm of faith and certainly are valid for people to embrace and believe if they choose. It is precisely because of this that faith cannot be disproved or otherwise assessed scientifically.

Much of the religiously based argument for the shroud's authenticity is based on the uniqueness of the Resurrection as told in the Gospels. The first major issue concerning this is that scholars do not all agree that the Resurrection in the Gospels, as interpreted today, is accurate. The Gospels were written well after the time of Jesus's life by at least a number of decades, and the story of a

physical Resurrection may have been a later interpretation. It is likely that for the Jews of Jesus's time, the idea of a resurrection meant something different, referring to an inclusion in the community as a whole at the final day of God. Debate on this issue remains, and many individuals would eschew this interpretation in deference to their own faith-based beliefs, making it a relatively ineffectual point of contention. Still, it is important to remember that the events in the biblical narrative may not be accurate records of real-life history. Perhaps the largest biblical challenged to the shroud's authenticity is that the text does not mention it at all. Instead, it mentions a pair of cloths and nothing of a miraculous image. Various theories have voiced concern whether the shroud was one of these cloths or an original temporary cloth, but they lack support from historical sources, functioning more as ad hoc arguments.

Argument Based on Radiation Theory

Those who support the idea that Jesus's Resurrection removed the body from the tomb and subsequently altered the shroud look to the realm of physics for their validation. They posit that the body of Jesus could have been transformed into energy in the form of radiation. Physics does allow for such a transformation, but not within the confines of the biblical stories. The energy released in such a transformation would be extremely vast, sufficient to destroy the shroud and much of the surrounding countryside. It is possible to fall back to a faith-based position, effectively that the transformation occurred but was somehow divinely limited. This position reenters the realm of nonscientific evidence, with all the problems inherent in such a stance. Looking past the problem of the limited radiation force, we again are forced to deal with the issue of image clarity. Radiation can be focused through various scientific means. The transition of a body into energy would possess no such ability to focus itself, resulting in a distorted, or likely, undistinguishable image. Looking at it as an effect similar to a camera, an analogy made by some shroud proponents, it would be more like exposing an entire role of film to the sun and then taking a picture through the camera lens. The shroud would be expected to show a blackened circle instead of a human image. As a scientific argument, the focused radiation image hypothesis is unsupported and contrary to observable evidence.

A further argument made, based on the theory that Jesus's Resurrection released a form of radiation, concerns the 1989 radiocarbon dating of the shroud. The hypothesis is that the radiation may have altered the chemical makeup of the carbon isotopes in the fabric of the shroud. This would potentially cause the material to date to a much younger period than it was actually from. Potentially, this could be true, but scholars and shroud proponents have no method to prove it. This theory does not follow the scientific assumption of uniformitarianism, meaning, that it is not based on observable events that have happened, been recorded,

measured, and studied. In short, scientists have never witnessed a divine resurrection and studied what it does to the radiocarbon age of fibers. While there is no way that the hypothesis of resurrection-altered carbon-14 cannot be disproved, neither can its supporters point to an event where this has been known to occur.

Conclusion

No matter what how intensely the Shroud of Turin is studied and found to be a 14th-century forgery, there will doubtlessly be some who continue to approach the question of the cloth's origin and nature from a faith-based position. There is nothing wrong with this, and for them the shroud will always be tied to the biblical story of Jesus's Resurrection. Using this belief as evidence to scientists, historians, and archaeologists remains impossible. These scholars work through methodology that does not function in the realm of faith, instead being limited by testability and uniformitarianism. Divine radiation and similar arguments that fail these requirements must then be left to people who already base their decisions on religious beliefs as to the origin of the shroud.

There are other explanations that have been formulated for the creation of the shroud in its many years of popularity. However, the evidence for such positions is often less robust than the theory of a painted shroud and is home to various holes of missing facts. One such theory is that the shroud was an early attempt at photography by Leonardo da Vinci. This is quite a puzzling idea, as the shroud clearly shows marks of paint and was known to exist before Leonardo was born. Not surprisingly, the cloth has also attracted tales of the Knights Templar. However, there is no historic evidence that the shroud was ever brought back from the Crusades by the knights. These fanciful histories are attractive in that they offer an explanation for the shroud's origin, which some consider hidden in mystery, but in the end, they ignore the scientific analysis of the shroud just as much as any explanation of the shroud being Jesus's burial cloth.

The concluding word on the Shroud of Turin must be this: it was not wrapped around Jesus in his tomb after he was crucified. There is quite a bit of evidence as to what actually occurred or to how the shroud came to be. While none of the evidence is, on its own, a definite point of proof, in conjunction they present a robust theory that cannot be ignored. The cloth was not known of until the 14th century, when it first went on display. Soon after this, an artist actually confessed to forging it through completely natural means. Modern analysis has come to support this interpretation. Radiocarbon dating places the shroud's creation at the time the confessor said it was. Historical and experimental studies have shown that the shroud could have been made during that period, even though it might appear strange to modern viewers. Finally, the presence of coloring agents on the shroud corresponding to the body and blood makes it quite clear that this was the method by which the image came to be.

Shroud supporters have attempted to argue against these points, finding potential areas of dissonance or questionable tests. By and large, these accusations have proven unlikely. It is also important to remember that even if something as important as the carbon-14 dating were found to be in error, this would not support the position that the shroud was originally used in the burial of Jesus. It is necessary that any such claims actually present positive proof, not just negative analysis of other studies. In the absence of such data, we must work from the null hypothesis that the Shroud of Turin was not used to bury the biblical Jesus.

References and Further Reading

Damon, P. E., et al. "Radiocarbon Dating of the Shroud of Turin." *Nature* 33 (February 16, 1989): 611–15.

Fagan, Garrett G., et al. *Archaeological Fantasies*. New York: Routledge, 2006.

Frei, Max "Nine Years of Palynological Studies on the Shroud," *Shroud Spectrum International,* 3 (June 1982), 3–7.

Gove, H. E. "Dating the Turin Shroud—An Assessment." *Radiocarbon* 32, 1 (1990): 87–92.

McCrone, W. C. "The Shroud of Turin: Blood or Artist's Pigment?" *Accounts of Chemical Research* 23 (1990): 77–83.

Nickell, Joe. *Inquest on the Shroud of Turin*. Buffalo, NY: Prometheus, 1983.

Nickell, Joe. *Looking for a Miracle: Weeping Icons, Relics, Stigmata, Visions and Healing*. Buffalo, NY: Prometheus, 1993.

Shermer, Michael. *Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of Our Time*. New York: Freeman, 1997.

7

A Staffordshire inscription points to the location of the Holy Grail; it may be in Wales.

PRO John Lee
CON Juliette Wood

PRO

The Holy Grail, substance of legends and myth, has been the subject of entertainment for countless generations of people and countries, especially Welsh and Scottish cultures. In their myths, legends, and stories, the Holy Grail has attracted the attention of scholars of literature and history to discover the ancient origins of this story. This legend has also provoked amateurs and professional scholars to discover the location of the Holy Grail. Several scholars have discovered inscriptions from Staffordshire, England, that point to the grail being located in Wales. This section attempts to support the claim of Staffordshire through the inscriptions and presents a discussion of an historian's interpretation of this claim.

The Holy Grail comes from the Latin *gradale* in the 19th century meaning dish or cup (Barber 2004: 95). It has also been associated with *krater* or the two handles on a cup or Greek vase. The word has been used in Catalonia in the wills and accounts of the people as two *gradales* or cups. In addition, it has been associated as a broad large dish, which was an object of value for the rich and the famous (Barber 2004: 96). The first use of the term grail as referring to the Holy Grail was in the medieval romances, with the origins of the word grail from the French culture.

The Legend of the Holy Grail

The Holy Grail was a cup, plate, or dish that Jesus Christ drank or ate out of during the Last Supper before his Crucifixion. Legend has it that Joseph of Arimathea, a Jewish Sadducee priest and a merchant, had it in his hand to catch the blood of Jesus Christ at the cross and had transported it to Wales, in England. It has also been claimed that Mary Magdalene, Jesus's female friend, had the cup. Also, some Mormon scholars have speculated that Mary Magdalene (along with other women, possibly) was married to Jesus Christ and that she produced heirs for him. The lineage of Christ was known as the sangreal or royal lineage. This forms the basis of the story of the Holy Grail, but there is an additional

125

Joseph of Arimathea collecting Christ's blood at the Crucifixion. From the Quest of the Holy Grail and the Morte d' Arthur, about 1300–1350. (The British Library/StockphotoPro)

twist that the Knights Templar had taken it to England on their journeys. This story can be found in Wales in the legends and mythologies of the Welsh people. These legends have placed Joseph of Arimathea as the central figures of the legends with Percival being the knight who saved the grail. The stories of King Arthur have the Knights of the Round Table searching for the grail in various places in Britain. These stories trace heroic episodes that the knights had to face trials before actually seeing the Holy Grail. The Cathars, a religious group, the Templars, a monastic knight group, and the Gnostics have been added to the legend as either carriers of the grail or part of Mary Magdalene's connection to the Holy Grail.

The First Theory of the Grail: Sangreal or the Royal Blood Line of Christ

The Holy Grail has also been symbolized as the holder or container of the royal bloodline or sangreal (blood royal) of Jesus Christ and the attendant seed of Jesus Christ. Mormon author Vern Grosvenor Swanson argues that Jesus Christ shed atoning blood to save the world, but he also donated his blood to a royal bloodline (Swanson 2006: 10). He proposes that Jesus was of the Judaic lineage,

while Mary Magdalene and Mary were of the tribe of Ephraim (45). According to Mormon beliefs, these two tribes were separated in the past but reunited in the royal marriage of both lines in Jesus and his wives. Jesus was the inheritor of the Israeli kingship through his bloodline from David and married Mary, who was an inheritor of the Israeli royalty through the tribe of Ephraim. Together they created a royal family. Because of this royal family, through children of Jesus, they created a lineage that ultimately ended in Joseph Smith, who became the inheritor of the kingdom of Jesus Christ (Swanson 2006: 344).

Through an elaborate tracing from before the beginning of time, Swanson traces how the lineage of Jesus Christ went to Great Britain where Jesus learned the secrets of the gospel through the Druids in Britain and brought back this knowledge to Jerusalem (Swanson 2006: 41). As part of this lineage, Jesus was destined to be king. Jesus goes there to learn Nicodemus's trade, who is his caretaker, learns about his genealogy and family, teaches his family about his truth, and establishes a church in Glastonbury and learns of the Druidic mysteries (Swanson 2006: 42–44). The Druids were religious intellectuals who studied the stars, math, architecture, and other universal secrets. Then through three cultural imperatives, Swanson claims that Jesus had to have been married. The first were the Gnostics, where in the Gospel of Thomas, Mary Magdalene is portrayed as the receiver of special knowledge and mysteries and who had a special relationship with the savior, which implies marriage (Swanson 2006: 55). Then he proves that marriage was sanctioned by Judaic law as an obligation of a rabbi and every Jewish man (Swanson 2006: 71). Then he proves that Jesus had to have been married, according to Mormon belief, in eternal marriage as was obligatory for Jesus as well as his followers (Swanson 2006: 78). He then traces Jesus's travels through the medieval legends by claiming that King Arthur might have been a descendant of Joseph of Arimathea (Swanson 2006: 215). He links the Cathars, who believed that Mary and Jesus had sexual intercourse, which the Catholics refute (Swanson 2006: 237). Then he weaves the Templars into the tale as the bearers of another secret lineage (a false lineage) that was protected by them, and then to attempting to find another secret in the temple (Swanson 2006: 241). He finally cites a third temple that was built by the Templars in the church at Roselyn and hypothesizes that the three kingdoms represent the two lineages united as one in the third column. Swanson returns to the British theory of their descent from Israel and the Hebrews and uses DNA tracing to show the descent from Jesus to Joseph Smith, who is the *messiah* who will restore the gospel. He claims that Joseph Smith knew that he was descendant from Jesus Christ and therefore a king of Israel.

The Second Theory of the Grail: Arthurian Romances

The Holy Grail has been discussed in British literature for many centuries with the first discussion by Chrétien de Troyes, an author of knight stories and the

author of *The Story of the Grail* (Barber 2004: 15). This story is about the encounter of Percival, who sees the grail in a procession. He has had to accomplish many trials to obtain even this vision. Chrétien writes three continuations that elaborate on this story by involving Sir Gawain (Barber 2004: 13–15). Chrétien conceives of the grail as a mystery and something that ignites awe in the beholder. He describes how Percival was confused by candles while seeing the grail (Barber 2004: 92). In addition, the grail is a holy thing that carries the Eucharistic sacrament. It is precious and made of gold (Barber 2004: 93). The grail is also a source of food, which supplies blood and wine to the Knights of the Round Table (Barber 2004: 101). In addition, Chrétien connects the grail to the Christian faith as the dish of the Lord's Supper (Barber 2004: 93–94). In addition, the grail is described as being used to heal Lancelot and other knights from wounds (Barber 2004: 101). Thus in Chrétien, the Holy Grail is a mysterious sacred object that provides miracles of healing, food, and spiritual nourishment for the knights.

The Holy Grail is next described by Robert de Baron, who wrote *L'Histoire du Graal* in the 1200. This story traces the history of the grail from the Gospels and places Joseph of Arimathea as the hero of the story (Barber 2004: 41). Joseph's brother-in-law, Bron, becomes one of the knights of King Arthur's Round Table. Arthur's father, Uther Pendragon, is the founder of the Round Table, but this is a parallel to the table of the Last Supper, the grail knighthood, and finally the Round Table itself. In the Round Table, there is one seat left vacant, which is supposed to be the seat of the future king of the Roman Empire, namely King Arthur himself (Barber 2004: 44). Finally, Percival is made keeper of the grail after Bron dies (Barber 2004: 45). This book is trying to link the Holy Grail with the Gospels and history. Furthermore, the grail symbolically becomes the dish or chalice of the Catholic Mass. For Catholics it is a symbol of the cup that collected Christ's blood and is intimately connected to the Mass (Barber 2004: 98). It becomes the ultimate symbol of Jesus Christ celebrating the Mass (Barber 2004: 98).

The third book is called *Perlesvaus* written by Jean de Nesle of Flanders in 1239. This book describes the character Perlesvaus as the assistant of Sir Lancelot as they trace the path of the grail to Arthur's castle (Barber 2004: 39). Perlesvaus attempts to defend the castle from 12 knights and he kills himself, but the Holy Grail appears and heals him (Barber 2004: 39). This story discusses Sir Lancelot's part in the grail legend.

The fourth book is *Lancelot*, which tells the story of Lancelot, Guinevere, and the ill-fated Camelot (Barber 2004: 57). The grail becomes a healer and a provider of food to these knights. The last grail book is by Wolfram von Eschenbach, a German writer around the turn of the twelfth century, who discusses the grail in the company of the angels in his *Parzival* (Barber 2004: 83).

In addition to the French origins of the grail, historians have proposed that it has derived from the legends of Celtic myths. Theodore de la Villemarque, in the *Conte populares Bretons* (1840), describes the grail in the bardic basins of

Celtic stories and legends (Barber 2004: 240). Villemarque believes that the Breton fables inspired the French romances with their ideals of the grail. Ernst Renan wrote *La poesie des race Celtiques*, which explains that the grail is a quest for sovereignty and an initiation of knights as contained in the poems of Peredeur's initiation (Barber 2004: 241). Thus the grail becomes the barding cups or initiation stories of the Celtic myths.

Richard Barber, the author of *The Holy Grail*, believes that the Holy Grail comprises the secret traditions of mystical practices suppressed by the Catholic Church (2004: 321). These practices were contained in a medieval book called *The Sworn Book*, which contained secret names and rituals that would allow the follower to obtain vision of Christ in 28 days. The rituals consisted of fasting, practicing Mass twice, praying, and reciting the prayers of the book (Barber 2004: 389). After performing these, the practitioner would obtain a vision or trance of Jesus Christ.

This is similar to the story where a knight, through self-discipline and denial as well as trials obtains, a vision of the Holy Grail (Barber 2004: 389). So the Holy Grail is this secret tradition that is hidden in *The Sworn Book*, and the Catholic Church suppressed the secret tradition because it threatened the ritual and the control of the priests by allowing individuals to achieve a vision by their own efforts in 28 days (Barber 2004: 389). These romances are for the select few who understand the symbolism of the Holy Grail. The Catholic ritual forced individuals to undergo the ardor of priesthood and recitation of the Mass for a longer period of time. These stories and legends have inspired the imagination and hope of many generations of readers. The Holy Grail has come to symbolize the perfection or the hopes of a generation and has been a symbol of the quest of the human soul for the divine. In recent years, the legend has become a target for conspiracy theories as contained in the book and movie *The DaVinci Code*. It has also become an expression for perfection and ardor in the newspapers and magazines. As a result, the Holy Grail has been a mysterious concept throughout Western literature and religion to symbolize the search for human perfection and imagination, which could be another origin of the grail.

The Third Theory: Staffordshire Location

The grail is placed in Staffordshire because of a stone called the Shepherd's Monument where an inscription is carved on the face. This inscription describes the lineage of Jesus Christ, who was supposed to have been entrusted to carry the grail unto the present day. This Shepherd's Monument is located in Shugborough Hall. The grail was also believed to have been hidden away in White Castle, which some historians believe is Whitton Borough Castle near the Shropshire border. It is at the Shropshire border that a cup that bears the name *grail* was discovered near a gravesite.

Shugborough Hall

One of the prominent sites in England related to grail lore is Shugborough Hall, the ancestral home of the Earls of Lichfield, in Staffordshire. Maintained today by the Staffordshire County Council and the British National Trust as a stately home open to public tours, Shugborough Hall is home to the Shepherd's Monument, which some believe points to the home of the Holy Grail. The monument, which was commissioned in 1748 by Thomas Anson, the earl at the time, is a marble slab depicting a group of shepherds examining a tomb, in a mirror image of Nicolas Poussin's painting "The Shepherds of Arcadia." The interesting part to grail lore is the seemingly haphazard arrangement of 10 Roman letters that many believe is a code pointing to the grail's location. Many have speculated on the code's meaning, and in 2004 two code breakers of World War II fame from England's noted Bletchley Park attempted to crack the code, but to no avail. However, in 2006 Louis Buff Parry, a Canadian cryptologist, claimed to have deduced some of the code's meanings, stating that the symbols indicated that the grail was buried somewhere close, although it has yet to be found.

The stone is hidden in a shepherd's monument built by Thomas Anson in 1748. In the monument, there is a marble tablet that is about 20 feet wide and 2 feet thick. The tablet has a picture carved in it of a shepherd and shepherdess contemplating heaven. The tablet has inscriptions on the bottom in two lines with 10 letters separated by periods. The stone inscriptions say D.O.U.O.S.V.A.V.V.M. An unknown hand inscribed these codes, but the builder, Thomas Anson, has links to the Prier de Sino, a secret society of Templar. Poussin, the French artists, drew an inverse image of the shepherd's stone that has the inscription on it. He has connections to the Prier de Sion (BBC, March 17, 2006). Some believe that Poussin inverted the letters as codes to the Templar (BBC, March 17, 2006).

Louis Buff Parry, a cryptologist, has attempted to translate the cryptic message. He says that the *D* and *M* stand for 1,500, the Roman numeral, which signifies the 1,500th verse of Genesis. The *VVA* stands for bloom or the bloom of Joseph that is contained in the 1,494th verse of Genesis. It believes the stone builder's stone has been lost. Parry believes that the stone will be found in Staffordshire and the Holy Grail will be discovered there (BBC, March 17, 2006).

Another cryptologist was an American code breaker who used a code matrix and discovered that if the letters *SEJ* were inversed they would create the word Jes or Jesus Defy. Historians believe that this was a symbol of a Christian sect that believed Jesus was an earthly prophet instead of the Son of God (BBC, November 26, 2004). He used this as a keyword for the rest of the code and discovered 1,2,2,3 as numbers in a code matrix on the side of the monument, which he believes will point to the location of the grail.

The location at White Castle also has shepherd's songs that might be linked to the one in Staffordshire (BBC, November 26, 2006). These songs denote the Arcadian or pastoral themes of the Staffordshire monument, and some historians believe this to be collaborating proof of the Staffordshire claim.

In conclusion, the Staffordshire claim is more believable because there is actual physical evidence claiming the existence of the grail. In addition, the Staffordshire claim has connections to actual people, who were connected with the Templar. In addition, there are two physical evidences—namely, Staffordshire and the White Castle location—that cite the shepherd song as part of the grail story. The other sources of the grail legend are based on speculation on royal lineage and romance stories, which are not as convincing as physical proof.

References and Further Reading

Baigent, Michael, Richard Leigh, and Henry Lincoln. *Holy Blood, Holy Grail*. New York: Dell, 1983.

Barber, Richard W. *The Holy Grail: Imagination and Belief*. Cambridge, MA: Harvard University Press, 2004.

British Broadcasting Company (BBC). "Code Points Away from Holy Grail." November 26, 2006. Available at http://news.bbc.co.uk/2/hi/uk_news/england/beds/bucks/herts/4040127.stm (accessed June 1, 2010).

British Broadcasting Company (BBC). "Holy Grail Lies 'at Stately Home.'" March 17, 2006. Available at http://news.bbc.co.uk/2/hi/uk_news/england/staffordshire/4818084.stm (accessed June 10, 2010).

Donington, Robert. *Wagner's "Ring" and Its Symbols: The Music and the Myth*. London: Faber, 1963.

Franke, Sylvie. *The Tree of Life and the Holy Grail*. East Sussex, UK: Temple Lodge Publishing, 2007.

Gardner, Lawrence. *Bloodline of the Holy Grail: The Hidden Lineage of Jesus Revealed*. Beverly, MA: Fairwinds, 2002.

Gardner, Lawrence. *The Magdalene Legacy: The Jesus and Mary Bloodline Conspiracy: Revelations beyond the Advance Code*. San Francisco: Weiser Books, 2007.

Glatz, Carol. "At Mass in Valencia, Pope Uses What Tradition Says Is Holy Grail." *Catholic News* (July 10, 2006).

Goering, Joseph. *The Virgin and the Grail: Origins of a Legend*. New Haven, CT: Yale University Press, 2005.

Hansen, H. T. "Foreword." In *The Mystery of the Grail: Initiation and Magic in the Quest for the Spirit*, by Julius Evola, Rochester, VT: Inner Traditions, 1996.

Loomis, Roger Sherman. *The Grail: From Celtic Myth to Christian Symbol*. Princeton, NJ: Princeton University Press, 1991.

Swanson, Vern Grosvenor. *The Dynasty of the Holy Grail: Mormonism Sacred Bloodline*. Springville, UT: Cedar Fort, 2006.

Wagner, Wilhelm. *Romance and Epics of Our Northern Ancestors, Norse, Celt and Teuton*. New York: Norroena Society Publisher, 1906.

CON

Legends regarding the location of the Holy Grail, the cup of Christ, with which Joseph of Arimathea is said to have caught the blood of Jesus at the Crucifixion, are many. The grail has been thought to be in Israel, Syria, France, England, and the United States, among many other locations. One of the most persistent grail legends has to do with a certain inscription on a monument in Staffordshire, England, which supposedly indicates that the grail is close-by. However, despite the fact that the inscription was made over 250 years ago, and intensive searches have taken place ever since, the location of the grail has never been determined. Despite the efforts of a legion of amateurs and many professional code breakers (such as the famed British cipher experts from Bletchley Park), nobody has ever been able to conclusively state exactly what the monument is communicating and even if it is regarding the grail. By looking through the history of the monument and its builder, we will be able to see that it is most unlikely that the encoded letters on the monument have anything to do with the grail.

Shugborough Hall is situated in Staffordshire, England, not far from the city of Lichfield. In the 17th century, a local lawyer, William Anson, purchased a house and some land, which became the core of an important estate. A new house was constructed to reflect the family's growing status, but the most significant change to the house and gardens was the work of the two great grandsons of the first William Anson, Thomas and George. Thomas, born in 1695, was a sophisticated, educated man with an interest in the classical arts and architecture of ancient Greece and Rome. His brother George, born in 1697, was a famous naval officer who sailed around the world between 1740–1744. During this eventful voyage, his ship captured a Spanish treasure galleon. The prize money made him immensely wealthy, and some of this wealth was used to improve the estate and the house. In 1747 he was named Lord Anson, and the following year he married Lady Elizabeth Yorke, daughter of the first Earl of Hardwick. The Ansons were a prominent family who wanted to create a setting befitting their wealth and prestige. Between 1745 and 1748, just before the new Lord Anson and his bride came to live at Shugborough Hall, the architect Thomas Wright carried out a number of extensions and improvements. Lord and Lady Anson had no children, and Thomas inherited his brother's fortune, which enabled him

to make more improvements to the house and grounds. Thomas Anson was a member of the British Parliament and active in the local community. He commissioned his friend, the architect James Stuart, to design a series of eight monuments for the garden and parklands. The landscaped garden at Shugborough Hall is typical of the period, containing monuments and decorative statuary as well as plants. These structures embody both the sophisticated aesthetics of the 18th century and the personal ideals of an influential and cultured family. Subsequently, the family suffered severe financial reverses, and a large proportion of the house contents were sold in 1842.

Today the house and the estate are open to the public. Information about the Anson family during the period when the Shepherd's Monument and other monuments in the garden were being constructed can be found in the family and estate papers. Many of these have been deposited in the Staffordshire Record Office and the William Salt Library in Stafford. The history of the monument on the Shugborough Hall Web site is based on these sources. One of the monuments in the garden, known as the Shepherd's Monument, has become associated with the mystery of the Holy Grail and a secret society of warrior knights. It has a carved bas-relief based on a painting by the French painter Nicolas Poussin. The painting depicts shepherds gazing at a tomb inscribed with the words "Et in Arcadia Ego." There is also a series of cryptic letters, $_D$.O.U.O.S.V.A.V.V.$_M$., with the first and last ones lower than the rest. This is surrounded by a rustic stone arch, which in turn is set inside an outer structure carved in the Doric Greek style. The letters and the fact that the carved relief is a mirror image of Poussin's original painting have attracted the attention of cryptologists and grail hunters for whom the letters present a code that reveals the true meaning of the Holy Grail.

Background

The story of the Holy Grail first appeared in medieval romances written in Europe in the 12th century. The grail was a mysterious object associated with abundance and danger. It was identified with the cup from which the founder of the Christian religion, Jesus Christ, drank at the Last Supper on the night before his death. On that occasion, he blessed bread and wine and shared it with his apostles. This event is still commemorated among Christians in Communion services and in the celebration of the Mass. Accounts from the New Testament in the Bible say that a man named Joseph of Arimathea offered to bury Christ in his own tomb. According to medieval romances, Joseph used the same cup to catch Christ's blood as he was being prepared for burial. This event is not mentioned in biblical accounts. The combination of biblical and legendary material forms the basis for the medieval story about the Holy Grail, which was used by Jesus Christ and brought by Joseph of Arimathea to Britain during the

time of King Arthur. The knights of King Arthur's court undertook a quest to find the Holy Grail, and a few worthy knights succeeded. After the quest was achieved, the grail was taken away into a supernatural realm and never seen again. Medieval romances were more concerned with the adventures of the knights than with the theological significance of an object associated with the sacrament of Communion and the Mass. However, since the revival of interest in Arthurian tradition in the 19th century, the grail itself has fascinated many writers. Numerous theories about its meaning, history, and present location have been put forward. In the context of the search for the true meaning and actual location of the Holy Grail, medieval romances are viewed not as fictional stories, but as codes that will lead to the discovery of a great secret. The knights who searched for the Holy Grail are equated with a real group of warrior knights known as the Knights Templar, or the Templars.

The Order of the Knights Templar was founded to defend Jerusalem and the Holy Land during the Crusades. When Jerusalem was finally lost, the Templars returned to Europe. They were wealthy and powerful and eventually they clashed with the king of France. At the beginning of the 14th century, the French king had the Templars in France arrested on false charges, claiming that they worshiped pagan idols and indulged in obscene practices. Although not everyone believed these charges, the Order of the Knights Templar was disbanded. When the accounts of the Templar trials were reexamined several centuries later, some writers suggested that the Templars were persecuted not because the French king wanted to destroy a rival institution, but because the Templars had learned some esoteric secret during the years they spent in the East and that this secret threatened the power of church and government alike. There is no clear evidence that the Templars were involved with esoteric matters, and very few historians have considered the possibility very seriously. However, the idea that the Templars have guarded a secret connected with the Holy Grail has been a mainstay of popular alternative history. For such popular historians history is one vast conspiracy to hide a secret. Certain events, like the suppression of the Templars, and objects, like

Shepherd's Monument at Shugborough Hall in Staffordshire, England. (Getty Images)

The Knights Templar, the Holy Grail, and Pre-Columbian America

The 14th century was not a good time to be a member of the Knights Templar, which had until then been the most wealthy and powerful of the military monastic orders created during the Crusades. At the behest of the King Phillip IV of France, Pope Clement V disbanded the order in 1312, and remnants of the order spent the rest of the century trying to elude capture and possible execution. Legend has it, however, that the Knights Templar possessed the Holy Grail, the cup that Joseph of Arimathea used to catch Jesus Christ's blood at the Crucifixion. Recently, some scholars have argued that the last of the Knights Templar left France in 1398 aboard their ships, sailed first to Scotland, then following the Viking voyages across the islands of the North Atlantic, finally settling in North America almost 100 years before Christopher Columbus sailed.

According to the legend, a Scottish prince named Henry Sinclair led a group of Knights Templar to Nova Scotia, where their presence is testified to in the mythology of the Micmac Indians of the region. Some archaeologists claim to have found geometric arrangements similar to those used by the Freemasons, who claim their heritage from the Knights Templar. But what might be even more interesting to some medieval enthusiasts is the possibility that they may have brought the Holy Grail with them and that it is today somewhere in North America.

Poussin's painting and the Shepherd's Monument, hold a key that will ultimately unravel this conspiracy, if researchers apply the correct methods. In relation to Shugborough and its supposed connection with the grail, the proposed key to the code resides in a secret history of the Templars (Baigent Leigh and Lincoln 1996), mysterious structure and lines in the landscape (Andrews and Schellenberger 2005), and psychic visions (Collins 2004).

Romantic ideas about warrior knights and hidden codes provide the context for the suggestion that the Anson family who lived at Shugborough Hall were somehow involved with a society of secret Templars who had survived the suppression of the order in the 14th century. The letters and the carving on the Shepherd's Monument have been interpreted as a code that will lead to the Templar's greatest secret, the location of the Holy Grail. References to the monument in contemporary documents, many of them in the personal documents and letters of the Anson family, reveal a great deal about the family and their attitudes toward art and life. They do not support the idea that the code has anything to do with secret societies or the Holy Grail. What these references do reveal is a newly wealthy family with sophisticated tastes who had access to the art, architecture, and literature of classical Europe. They drew inspiration for how they wished to live their own lives from the classical Greek and Roman world, and they expressed these aspirations in the architecture and garden design of their home at Shugborough.

History

Three designers were involved in the construction of the Shepherd's Monument at Shugborough Hall. Thomas Wright of Durham, an architect, garden designer, mathematician, and astronomer, was employed by the Anson family to extend and improve the house, and he also drew the original design for the monument at Shugborough. The Poussin relief was executed by the Antwerp-born sculptor Peter Scheemakers from a print of the French painting by Bernard Picart. The Doric-style surround was added later by another architect designer, James Stuart, who was a friend of the owner and an important figure in the revival of interest in classical architecture and culture during the 18th century.

The origins of this so-called grail mystery, however, are not rooted in the 18th-century world of the Anson family. The grail mystery is linked to the 20th-century world of a group of French grail enthusiasts who created an organization called the Priory of Sion. They claimed that the Priory formed the inner core of the Knights Templar and guarded their most precious secret in order to provide an imaginative, but completely synthetic, history for this secret society. These modern enthusiasts produced a number of mysterious documents. They interpreted details in Poussin's painting, such as the tomb with its inscription and the figures who appeared to point to features in the landscape, as references to an actual place in southern France, supposedly the secret last resting place of the Holy Grail. However, neither the tomb nor the landscape in Poussin's painting reflects real geographic features. Poussin used this imaginary tomb inscribed with the words "Et in Arcadia Ego" in several paintings as a symbol for mortality and the transience of life. Although a structure was built several centuries later in the French countryside, no tomb existed when Poussin painted the Arcadian Shepherds in the 17th century, and it is unlikely that he ever visited this part of France (Putman and Wood 2005: 115–32). However, the authors of the most popular alternative history about the Holy Grail inserted the Shugborough monument, with its seemingly mysterious series of letters, into their speculative history about the French painting and the Priory of Sion (Baigent, Leigh, and Lincoln 1996: 190–91). Since then, a location with no prior links to the Templars or the Holy Grail has been absorbed into a dynamic, but unsupported, modern legend, and it has attracted new motifs of its own.

Not far from Shugborough is Bletchley Park, another country house in the midst of an estate. Code breakers at Bletchley Park, who worked for British intelligence during World War II, broke the infamous German *Enigma* codes. More than half a century later, the name still conjures up visions of secret agents, wartime espionage, and the fight against repression. Several new solutions to the meaning of the Shugborough monument were presented to the world media through the efforts of two code breakers who worked at Bletchley Park during World War II. This revived the wartime nostalgia associated with

the place and provided a seemingly authentic source for decoding the secret (Shugborough Estate Web site; *The Guardian* November 26, 2004, 3; *The Times* November 26, 2004; *The Daily Telegraph* November 26, 2004).

The most dramatic solution, however, did not come from Bletchley Park but from an unnamed code breaker working in a secret intelligence organization. This solution explained the letters as a reference to the Holy Grail and to the belief that the Templars allegedly preserved an alternative religious tradition in which Jesus Christ was human, not divine. The letters were submitted to a series of code grids to yield a "solution," *Jesus H Defy*. This makes little sense as it stands and in no way solves the puzzle. Further interpretation identified the *H* with the Greek letter chi (*X*), and translated the *X* as *messiah/Christ*. The phrase is thus explained as *Jesus (the Deity) Defy*. The reasons for these changes have never been fully explained. The Greek *H* consistently refers to the letter *e* in the name Jesus, as in the abbreviation IHS, the first three letters of the name in Greek. The words *messiah* and *Christ* mean "the anointed one" not "deity." So, the code breaker adds more speculative history claiming that the Templars practiced an alternative Christianity that denied the divinity of Jesus. Unfortunately this has no more inherent sense than the original "code" and owes more to romantic ideas about the Templars and the popularity of alternative history than to any serious attempt to solve a code. In fact this is a circular argument. It asserts that the Templars practiced an alternative Christianity, without any concrete proof that this was so, uses this to interpret the "code," and then presents the solution as proof of the original assertion. References to code-breaking grids, an influential, unnamed code breaker from an intelligence network, the Templars, and the Priory of Sion are elements that make the contemporary grail legend so compelling. It is also these very elements that locate this explanation in the world of modern legend rather than history (Wood 2003). Indeed a reference to the "denial of Jesus' divinity" occurs in the paragraph immediately after the description of the Shugborough Hall code in the alternative history book where is was featured originally, so no code breaker need look very far (Baigent et al. 1996: 192).

Other solutions have been suggested for these cryptic letters besides a grail code. The 10 letters are separated by full stops, which implies that they are abbreviations for words. Several solutions offer Latin or English phrases using these letters. Margaret, countess of Lichfield, a member of the family presently occupying Shugborough, remembered a story she had heard as a child and proposed the following solution: "Out of your own sweet vale Alicia vanish vanity twixt Deity and Man, thou Shepherdess the way" (Shugborough Academy Web site). Although no trace of the story has ever been found, the explanation hints at the notion of vanity, and by extension, to a philosophy known as stoicism. This philosophy first appeared during the Hellenistic period (ca. third century BCE) and was popular among the educated Greco-Roman elite. Stoicism stressed that life and its blessings were transitory; therefore, a virtuous life was the basis

of true happiness. The stoics belonged to the classical world that educated 18th-century men and women admired. Such sentiments would have appealed to members of the Anson family. Another solution also echoes the notion of life's brevity. *"Orator Omnia Sunt Vanitas Ait Vanitas Vanitatum."* This Latin phrase paraphrases a biblical verse in Ecclesiastes, "Vanity of vanities, saith the preacher, all is vanity" (*Billings Gazette* June 6, 2006). Both of these solutions echo the sentiments of virtuous living and endurance, which those in the 18th century attributed to the classical world. Either solution could be seen as a comment on a scene depicting the perfect pastoral world of Arcadia whose beauty is disrupted by death.

It is also possible that the letters are a personal memorial to the memory of a departed loved one. The first and last letters, *D. M.*, were a standard abbreviation for *Diis Manibus* ("To the Souls of the Departed") and were carved on Roman funerary monuments. The remaining letters might stand for the Latin phrase "*Optimae Uxoris, Optimae Sororis, Viduus Amantissimus Vovit Virtutibus.*" This could be translated as "Best wife, best sister, the most loving widower dedicates [this] to [your] virtue." This is the most personal of the solutions offered. Once again the sentiments fit the poetic ambiguities of Arcadian symbolism popular with the Anson family and their circle. Depending on the date on which the inscription was carved, this could be a memorial to the parents of the Anson brothers who carried out improvements to the estate and gardens, or to Elizabeth, Lady Anson, who died in 1760. The cryptic letters may commemorate the affection between Lady Elizabeth (the wife and sister) and her husband (widower) who survived her by only a few years, or perhaps it may refer to the Anson's mother and father. It is even possible that the memorial commemorates an early and brief marriage of Thomas. Local records note the marriage of a "Thomas Ansin" to Anne Ridell in 1728. Although the exact identities of the wife and widower are still unclear, it does provide a solution to the code that becomes clear once it is "cracked" (Shugborough Academy, cited April 30, 2007).

If the Templars and the Holy Grail lie behind the monument and its mysterious lettering, then there should be some indication in the family papers or other documents. Secret codes are all too easy to manufacture if details are strung together independent of their historical and cultural contexts. The earliest element of the monument was based on a design by Thomas Wright (1711–1786), a mathematician, architect, antiquarian, and astronomer. The original design of a rustic arch on which this monument is based appeared in one of his books on architecture, which included a number of sketches for architectural features intended to adorn the houses of wealthy aristocrats. He described this collection of rustic follies and hermitages in terms of the fashionable images for past wisdom as "suitable for a Brahmin or a druid" (Harris 1979: plate A). Thomas Wright's interests mirrored those of his 18th-century patrons. He was erudite, elitist, intellectually playful, confident about the harmony of knowledge

and creation, and stoical about the vicissitudes of fate. At first glance the imaginative Thomas Wright seems an ideal purveyor of codes and secrets. His scientific fame rests on his explanation of Earth's position in the Milky Way galaxy, but his attempt to produce a cosmology integrating divine, moral, and scientific views was full of unusual notions. He also wrote utopian fiction. He was a Freemason and he used codes and ciphers in the decoration of his own home. When he was young, Wright's father, thinking his son was mad, burned all of the young man's books. He was an antiquarian with a particular interest in the beliefs of the Druids. He made drawings of ancient stone monuments in Ireland that he believed had been built by the Druids. Eighteenth-century ideas about what the Druids believed were an important source for his fantasies about the past and the inspiration for many of his designs. There is no mention of Templars in his mystical worldview. In any event, Wright was not involved in the creation of the bas-relief whose imagery and inscription have suggested secret meanings to some observers. In the finished monument at Shugborough, the mirror image adaptation of Poussin's painting of shepherds near a tomb in Arcadia set within Thomas Wright's rustic arch was executed by the sculptor Peter Sheemaker. Another architect, James Stuart, designed and built several other monuments in the garden. He also surrounded both the carving and Wright's rustic arch with a portico in the Greek Doric style.

It is the Ansons, the owners of the estate, and how they fit into the cultural interests of the 18th-century intelligentsia that can tell us the most about Shepherd's Monument. Admiral George Anson, famous for his circumnavigation of the globe and newly enriched from sea booty, came to live at Shugborough with his wife, Lady Elizabeth Yorke, whom he married in 1748. His brother, Thomas Anson, was the owner of Shugborough. He was a member of the Royal Society and a founding member of the Dilettanti Society, a dining club devoted to the revival of classical art. Both of these organizations were devoted to intellectual pursuits and were not secret societies. Thomas's friend, James Stuart (1713–1788), who did much to revive the Greek style as an architectural fashion, added the Doric surround to the Shepherd's Monument.

The significance of a tomb in Arcadia is important in order to understand the meaning of the monument. The question to be resolved is whether it conceals an esoteric secret or whether it had personal meaning for the Anson family. There are indications that the Anson family thought of the estate in Arcadian terms. Arcadia was a region in ancient Greece devoted to farming and agriculture. For this reason it came to symbolize the virtues of the pastoral life, one that was simple, untouched by ambition or corruption, and unchanging. Educated men and women of the 18th century saw themselves as the inheritors of Greek and Roman values and began to remake their environment in the image of the Greek and Roman world. Collecting classical antiquities became fashionable, and gardens were designed to imitate a romanticized vision of the pastoral simplicity of

Arcadia. Paintings of classical subjects also became popular. The Duke of Devonshire, an acquaintance of the Anson family, owned an earlier version of the Arcadian shepherds looking at a tomb painted by Nicholas Poussin. He lent this painting to Elizabeth, Lady Anson, who made a copy, and there is a picture of her holding the sketch still in the Lichfield family collection. Elizabeth Anson was a talented woman, well read in classical literature. She described Shugborough as "Arcady" in a letter to her brother-in law, Thomas, whom she also addressed as "shepherd" (Harris 2006: 1–2).

The Roman poet Virgil wrote a series of poems called *Eclogues* set in an idealized Arcadia and praised the virtues of a pastoral life. The phrase "Et in Arcadia Ego" echoes a passage in one of these poems, and it appears in paintings by later artists with interests in classical myth. The contrast in Poussin's *Les Bergers D'Arcadie* (*The Shepherds of Arcadia*) between the pastoral scene and the inscribed tomb suggests the clever ambiguity so beloved of sophisticated painters and their classically educated patrons (Blunt 1996). Their wealth enabled them to indulge in the study of classical literature and extensive travel, and many of them amassed collections of classical art and paintings from contemporary artists that incorporated mythic references. They also wanted to emulate the ideals of the Greek and Roman world in their own lives. The phrase "Et in Arcadia Ego" is deliberately ambiguous. It refers both to the transience of life, "even in Arcadia am I (i.e., Death)" and to the beauty of eternity, "I (the occupant of the tomb) am in Arcadia." There is another detail that may link the more general taste for Arcadian symbolism with personal meanings for the Anson family. The phrase was inscribed on a funeral urn commemorating the death of Henry Pelham, a close friend and political ally. An urn was carved above the tomb in the Shepherd's Monument at Shugborough, although no such object appears in the Poussin painting. This may be another personal reference in the monument to the death of a friend. Another similarity between the Pelham and Shugborough memorials is the phrase *optimae uxoris*. It appears on both monuments and, like *diis manibus*, it is also found on Roman grave memorials.

A poem written in 1758 about the monument at Shugborough titled "Hermit Poem on an Emblematical Basso Relievo after a famous picture of Nicolas Poussin" mentions Arcadia and the fact that "life's fleeting moments gently steal away." The subject of this poem is the carving of Poussin's work, but there is nothing mentioned about the cryptic letters. Another poem, written in 1767, calls them "mystic ciphers," but does not give any indication of what they might mean. During that period both Lady Elizabeth and her husband had died and Thomas Anson had engaged his friend, the architect James Stuart, to make further additions to the park and garden. The first volume of Stuart's important book on architecture, *The Antiquities of Athens*, appeared in 1762, the year before he started working at Shugborough. One of the drawings included in the volume echoes some of the elements in the Shugborough inscription. It

depicts a funereal shield with the letters *D.M.* and a Latin inscription (Stuart 1762). These classical references would have appealed to, and to some extent flattered, the accomplishments of the family and their circle of friends. In 1782 another friend, Thomas Pennant, came to visit Shugborough. He was a famous traveler who wrote an account of his travels through the British Isles at a time when such journeys were an impressive undertaking. He described both the gardens and Thomas Anson's attitude to them:

> The scene is laid in Arcadia. Two lovers expressed in elegant pastoral figures appear attentive to an ancient shepherd who reads to them an inscription on a tomb "Et in Arcadia" the moral resulting from this seems to be that there are no situations of this life so delicious but which death must at length snatch us from. It was placed here by the owner as a memento of the certainty of that event perhaps as a secret memorial of some loss of a tender nature in his early years for he was wont often to gaze on it in affection and fine meditation. (Pennant 1782)

This reference also describes the carving rather than the cryptic letters, but the author knew Thomas Anson personally. Pennant's description emphasizes the personal nature of the imagery and its connection with a stoic endurance of loss.

Conclusion

The announcement that the code had been cracked produced a flurry of interest, and, although this died down somewhat, the Shugborough Web site still contains a "Holy Grail" section. However, none of the details drawn from contemporary documents links the Shugborough monument to the grail or the Templars. What information there is about the interests of the Anson family places the monument firmly within the popular theme of a romantic, elegiac Arcadia, something that was widely known and appreciated in the 18th century. It is now clear, indeed it has been clear for some time, that Poussin painted an imaginary tomb. It was a symbol for mortality and he used it in several paintings. A structure was built on a site in the French countryside much later, and the link between the two was only made in the 20th century. Since the inscribed tomb did not exist when Poussin painted his vision of shepherds in Arcadia, there can be no connection with Shugborough or the family who commissioned the monument incorporating a version of his painting in the 18th century. Similarly the reversal of the composition, the urn, and the changed angles cannot be attributed to that fact that "Staffordshire was a hotbed of Masonic activity" (Baigent et al. 1996: 191). Nor can the figures and the lettering be interpreted as symbols pointing to "the location of the treasure—the tomb of god the holy blood and the holy grail" (Andrews and Shellenberger 2005: 88). The engraving used by Sheemakers was printed in reverse, and the broad rectangular composition of Poussin's landscape

had to be compressed into a narrower, "portrait" frame. Aristocratic families, like the Ansons, especially when they possessed something that appears to be mysterious, attract just this sort of legend.

The information that we have about the Ansons and their monument presents a very different picture from the rather wild speculations that Admiral Anson could have captured the bas-relief based on Poussin's painting at sea from a Templar ship. The connections between the Shepherd's Monument and the Holy Grail are tenuous at best, requiring a sizable stretch of the historical imagination. Thus, the mystery at Shugborough is not in any way related to medieval romance about the Holy Grail, nor does it really concern any grail relic. Attempts to create and then solve a mystery focus on the meaning of the mysterious cipher and its possible link to esoteric ideas. The Shepherd's Monument at Shugborough Hall seems to have had a personal significance for the family. The exact nature of that significance remains unclear, but a likely explanation is that it commemorates the loss of some family member using the imagery of Arcadia. By examining the existing sources, many of them contemporary, with the construction of the monument, we can, however, understand the cultural and social context in which the Shepherd's Monument was constructed, revealing a much more conventional meaning than grail seekers might like to see.

References and Further Reading

Andrews, Richard, and Paul Shellenberger. *The Tomb of God, the Body of Jesus and the Solution to a 2000 Year Old Mystery.* Rev. ed. London: Little Brown, 2005.

Baigent, Michael, Richard Leigh, and Henry Lincoln. *The Holy Blood and the Holy Grail.* London: Corgi, 1996.

Baker, Andrew, "The Shepherdess's Secret." Unpublished manuscript at the Staffordshire Record Office.

Billings Gazette. "City Lights: Parmly Plus Pluck Solve Old Puzzle." Available at http://www.billingsgazette.com/news/local/article_7b9cd7d5-568d-5260-b42a-68dbb27aad71.html (accessed August 3, 2010).

Blunt, Anthony. *The Paintings of Nicolas Poussin.* London: Phaidon, 1996.

British Broadcasting Company (BBC). "Code Points Away from Holy Grail." Available at http://news.bbc.co.uk/1/hi/england/beds/bucks/herts/4040127.stm (accessed June 1, 2010).

Collins, Andrew. *Twenty-First Century Grail: The Quest for a Legend.* London: Virgin Books, 2004.

Harris, Eileen. "Cracking the Poussin Code: The Key to the Shepherd's Monument at Shugborough." *Apollo* (May 2006). Available through title search at www.archives.staffordshire.gov.uk (accessed June 1, 2010).

Harris, Eileen, ed. *Thomas Wright's Arbours and Grottos: A Facsimile. With a catalogue of Wright's works in architecture and garden design.* London: Scolar, 1979.

Morris, Steven. "Has the Mystery of the Holy Grail been solved?" *The Guardian* (November 26, 2004) Available at http://www.guardian.co.uk/uk/2004/nov/26/artsandhumanities.highereducation (Accessed August 3, 2010)

Pennant, Thomas. *Journey from Chester to London*: Chester, 1782.

Putnam, Bill, and John Edwin Wood. *The Treasure of Rennes-Le-Chateau: A Mystery Solved.* Rev. ed. Stroud: Tempus, 2005.

Shugborough Estate. Home page. Available at www.shugborough.org.uk/AcademyHome-156 (accessed June 1, 2010). Shugborough Academy lists the history of the monument and selections from newspaper articles:

Smith, Lewis. "War Codebreaker Cracks an Enigma of Love." *The Times* (November 26, 2004) Available at http://www.timesonline.co.uk/tol/news/uk/article395668.ece, (Accessed August 3, 2010)

Stuart, James, and Nicholas Revett. *The Antiquities of Athens*. 1762. Vol. 1. London: Princeton Architectural Press, 2007.

Tweedie, Neil. "Letters Remain the Holy Grail to Code-Breakers." *The Daily Telegraph* (November 26, 2004)Available at http://www.telegraph.co.uk/news/uknews/1477527/Letters-remain-the-holy-grail-to-code-breakers.html (Accessed August 3, 2010)

Wood, Juliette. "The Templars, the Grail and Just About Everything Else: Contemporary Legends in the Media." *FLS News: The Newsletter of The Folklore Society* 45:2 (2003).

8

Nestorius did not intend to argue that Christ had a dual nature, but that view became labeled Nestorianism.

PRO   Mark Dickens

CON   Annette Morrow

PRO

History is rarely kind to heretics, even less so to heresiarchs, those who devise systems of belief that lead the faithful astray. Their stories are usually told not by themselves, but by their opponents. In the process, they are condemned for questioning the religious *status quo* or offering innovative solutions to theological problems. Their beliefs are minutely scrutinized to discredit their views, and they are often accused of immoral behavior, as further evidence of their heretical thinking. Since those who triumph over the heretics often destroy most or all of their works, one can only evaluate them through the lens of their opponents.

Most heresiarchs are universally regarded by scholars as clearly opposed to the basics of the Christian faith as outlined in the Bible and interpreted by the church leaders since apostolic times. However, about Nestorius there is much less consensus; for the past century, theologians have held widely divergent views on his teachings. Was he truly a heretic or rather a victim of church politics whose views have been subsequently misinterpreted, in part due to the exalted status of his opponent, Cyril of Alexandria? This section proposes the latter view; Nestorius does not deserve to be labeled a heretic because he did not teach what he is accused of.

Before going further, the idea of a dual nature needs to be clarified. The word is misleading, since Nestorius undeniably argued that Christ had a dual nature. This position, known as Dyophysitism (from Greek *dyophysitai*, "two natures," referring to the divine and human natures of Christ), is also the orthodox Christian position articulated at the Council of Chalcedon in 451 and considered a primary article of faith by the Catholic, Orthodox, and Protestant churches today. It stands in contrast to the Miaphysitism (from Greek *mia physis*, "one nature") of the Oriental orthodox churches (Coptic, Ethiopian, Syrian, and Armenian Orthodox), who affirm only one nature in Christ. By contrast, Nestorius is accused of teaching that there were two *persons*, not two *natures*, in Christ (a crucial

145

terminological distinction). This section, therefore, disputes the accusation that Nestorius taught "two persons in Christ."

Historical and Theological Background

After three centuries of surviving as an illegal religious sect within the Roman Empire, Christianity experienced a dramatic turnaround when Constantine I (306–337) issued the Edict of Milan (313), signaling the end of official state opposition to the faith. The Christianization of the empire proceeded apace over the next several decades, resulting in the proclamation of Christianity as the official state religion in 380 by Theodosius I (378–395).

The reprieve from persecution and subsequent state sponsorship of Christianity meant church leaders could turn their attention to unresolved theological issues that had been brewing for decades. There were both religious and political reasons for doing so. Church and state were increasingly interconnected, and most emperors viewed the ecclesiastical unity of the empire as inextricably linked to its political unity; solving theological problems had serious implications for governing the empire. Hence, beginning with Constantine I at the Council of Nicaea (325), emperors periodically convened ecumenical councils at which the gathered bishops debated issues vital to the doctrinal unity of the church.

The chief concerns at the first several ecumenical councils centered on two Christological issues: the relationship between the Son (Jesus) and the Father (God) in the Trinity and the relationship between divinity and humanity in Christ. The first concern was at the heart of the Arian controversy, which was addressed at the Ecumenical Councils of Nicaea (325) and Constantinople (381). The second issue, which was the core of the Nestorian and Monophysite controversies, dealt with at the Ecumenical Councils of Ephesus (431) and Chalcedon (451).

The essential problem is that the New Testament affirms both the divinity of Christ and his humanity, but does not clearly explain how the two interact with or relate to each other. Various biblical statements on this relationship can be interpreted in several different ways, notably John 1:14: "The Word became flesh and made his dwelling among us. We have seen his glory, the glory of the One and Only, who came from the Father, full of grace and truth" and Philippians 2:5–7: "Christ Jesus, who, being in very nature God, did not consider equality with God something to be grasped, but made himself nothing, taking the very nature of a servant, being made in human likeness."

Two contrasting views on the divine–human interaction in Christ were developed in the theological schools of Antioch (Syria) and Alexandria (Egypt). The Antiochenes followed a literal and historical approach to biblical exegesis, while the Alexandrians favored an allegorical and philosophical approach. The emphasis

that Antioch placed on the historical facts of Jesus's life resulted in a strong focus on his humanity, whereas the more metaphysical approach of Alexandria produced a greater emphasis on his divinity. Important representatives of the Antiochene tradition include Paul of Samosata, Diodore of Tarsus, Theodore of Mopsuestia, John Chrysostom, and Nestorius. The Alexandrian school produced Origen, Athanasius, Apollinarius of Laodicea, Cyril of Alexandria, and Eutyches (both lists contain saints and heretics).

A corollary of these different theological emphases was their approach to the role of the logos, the preexistent Word of God that became incarnate in Jesus Christ. Antiochenes generally spoke of the logos dwelling alongside the human in Jesus, resulting in two logical subjects in Christ (called *logos-anthropos* or "word-man" theology). In contrast, Alexandrians described the Word taking on flesh to such an extent that it became the sole logical subject of the person of Christ, with the practical result that his divinity often eclipsed his humanity in their thinking (called *logos-sarx* or "word-flesh" theology). The union was one of essence or substance in which the human will was eclipsed by the divine will. All Christological statements in the Bible, including those about Jesus's birth, suffering, and death, were ultimately ascribed to the divine logos become flesh (and therefore to God).

Taken to their logical extremes, both viewpoints could end up in heretical thinking. The Antiochene Paul of Samosata (d. 275) taught that Jesus was merely a man in whom the Holy Spirit dwelt; whereas the Alexandrian Apollinarius of Laodicea (d. ca. 390) argued that the human mind in Christ had been replaced by the divine mind of the logos. Apollinarius also coined the phrase "one incarnate nature of the God Logos," later used by Cyril, who thought it came from Athanasius, the great champion of Nicene Christianity. The teachings of both Paul and Apollinarius were subsequently condemned by church councils. On the same basis, many scholars would also include as examples of the heretical potential in the two competing theological systems the names of Nestorius and Eutyches (representing Antioch and Alexandria, respectively), both condemned by the Council of Chalcedon (451).

Another key factor was the increasing rivalry between the apostolic sees (or patriarchates) of Antioch: Alexandria and Constantinople. Initially, there had been near equality among Rome, Antioch, and Alexandria, with Rome being accorded the status of "first among equals." However, Canon 3 of the Council of Constantinople (381) moved the new capital Constantinople into second place after Rome (a position strengthened by Canon 28 of the later Council of Chalcedon). This move particularly irked the patriarchs of Alexandria, who looked back to the evangelist Mark as their apostolic founder and had called themselves "popes" since the patriarchate of Heraclas (232–248). By contrast, any claim by Constantinople to apostolic foundation had to be fabricated (and was, in the person of the apostle Andrew).

The Rise and Fall of Nestorius

When Sisinnius I, patriarch of Constantinople, died in December 427, Emperor Theodosius II (408–450) chose Nestorius, a Syrian monk and disciple of Theodore of Mopsuestia, to replace him. Nestorius was consecrated as bishop of Constantinople in April 428. The church historian Socrates Scholasticus describes him as "distinguished for his excellent voice and fluency of speech," but his subsequent actions revealed the "violent and vainglorious temperament," of one who "continually disturbed the public tranquility" (Stevenson and Frend 1989: 287–88).

Whether or not Nestorius was as arrogant as Socrates claims, his actions reveal why the people of the capital nicknamed him the "incendiary" bishop. In his inaugural sermon he asked the emperor's assistance in purging the realm of heretics. When he attempted to impose his authority over the Arians in Constantinople, a fire and riot ensued in the city. Demonstrating both religious zeal and political naivety, Nestorius proceeded to attack immorality in public entertainment, to bring the city's monks under his ecclesiastical jurisdiction, to restrict the involvement of aristocratic women in ecclesiastical affairs, and to challenge the role of the Augusta (empress) Pulcheria, the powerful sister of Theodosius II. In so doing, he alienated the general population, the monks, the aristocracy, and the empress. This would haunt him during the subsequent theological controversy; while his opponent Cyril of Alexandria (412–444) "had an immensely strong personal power-base in his own church. . . . Nestorius had set almost everyone against him on the home front" (McGuckin 1996: 20).

Although these actions played a role in Nestorius's eventual downfall, the main complaints about him concerned his Antiochene Christological views and particularly his rejection of the term *Theotokos*, "Bearer/Mother of God," to describe the Virgin Mary. When the presbyter Anastasius preached against the use of *Theotokos*, saying "It is impossible that God should be born of a human being," Nestorius backed him up and began to also preach against the term, urging instead the use of *Christotokos*, "Bearer/Mother of Christ," since it avoided the implication that divinity had its source in humanity. However, as Socrates notes, he "acquired the

Nestorius, Persian prelate and Patriarch of Constantinople who was deposed for his heretical views about the nature of Jesus Christ. (Mary Evans Picture Library/The Image Works)

reputation among the masses of asserting that the Lord was a mere man." Although Socrates concluded this was not what Nestorius actually taught, he critiqued him for not paying adequate attention to earlier theologians' use of *Theotokos* (Stevenson and Frend 1989: 288–89). Nestorius's rejection of the term was seen by the general populace, who increasingly venerated Mary, as an assault on their religious devotional life, and Pulcheria (a consecrated virgin with a reputation for prayer and good works) probably interpreted Nestorius's opposition to *Theotokos* as a personal attack (Russell 2000: 32–33). Meanwhile, to Cyril of Alexandria, Nestorius's position amounted to questioning the divinity of Christ.

Cyril wrote three letters, in increasingly urgent terms, attempting to bring Nestorius into line with his thinking on Christ's nature, but Nestorius stood firm. The dispute between the two came to a head with the Ecumenical Council that was called by the Emperor Theodosius, to meet at Ephesus on Pentecost in 431. Although Nestorius welcomed this as an opportunity to confront Cyril, the council was to prove his undoing, given the way his teaching and actions as patriarch had alienated so many. In anticipation of the council, Cyril wrote 12 anathemas, or accusations, which he attached to his final letter to Nestorius (Stevenson and Frend 1989: 307–8). To avoid excommunication, Nestorius had to agree to all 12 accusations. In issuing the anathemas, Cyril had clearly overstepped the authority delegated to him by Celestine. They presented a strong Alexandrian position to which no Antiochene could agree. Moreover, the dispute between Cyril and Nestorius had become so personal that the latter was probably beyond agreeing to anything the former proposed, even where there were grounds for genuine theological agreement.

The Council of Ephesus

In early June 431, approximately 200 bishops gathered in Ephesus. Apart from 10 who accompanied Nestorius, most were Cyril's allies, since John of Antioch and his delegation of 43 bishops had been delayed. Despite receiving a letter from John announcing their imminent arrival, the council began without them, under Cyril's leadership. Ignoring protests from the emperor's representatives and gathering in the Great Church of St. Mary the *Theotokos* in Ephesus, the council proceeded to depose and excommunicate Nestorius on June 22. Although summoned to appear, Nestorius refused, rightly understanding that he would not get a fair trial. When John and the Antiochene bishops arrived on June 26, they convened an alternate council and immediately deposed Cyril and Memnon, bishop of Ephesus and Cyril's ally, as well as excommunicated their supporters who refused to "anathematize the heretical propositions of Cyril" (Stevenson and Frend 1989: 309). When the pope's legates reached Ephesus on July 10, they supported Cyril, giving papal assent to Nestorius's deposition.

Nestorius's Letter to Cyril of Alexandria

At the Council of Ephesus, evidence against Nestorius, in the form of his correspondence with Cyril of Alexandria, was read. Ironically, it remains one of the only extant pieces of Nestorius's writing, as most others were destroyed after he was declared a heretic. In this segment, Nestorius explains his views on Jesus's nature:

Holy scripture, wherever it recalls the Lord's economy, speaks of the birth and suffering not of the godhead but of the humanity of Christ, so that the holy virgin is more accurately termed mother of Christ than mother of God. Hear these words that the gospels proclaim: 'The book of the generation of Jesus Christ, son of David, son of Abraham." It is clear that God the Word was not the son of David. Listen to another witness if you will: 'Jacob begat Joseph, the husband of Mary, of whom was born Jesus, who is called the Christ." Consider a further piece of evidence: 'Now the birth of Jesus Christ took place in this way. When his mother Mary had been betrothed to Joseph, she was found to be with child of the holy Spirit." But who would ever consider that the godhead of the only begotten was a creature of the Spirit? Why do we need to mention: 'the mother of Jesus was there"? And again what of: 'with Mary the mother of Jesus"; or 'that which is conceived in her is of the holy Spirit"; and 'Take the child and his mother and flee to Egypt"; and 'concerning his Son, who was born of the seed of David according to the flesh"? Again, scripture says when speaking of his passion: 'God sending his own Son in the likeness of sinful flesh and for sin, he condemned sin in the flesh"; and again 'Christ died for our sins" and 'Christ having suffered in the flesh"; and 'This is," not 'my godhead," but 'my body, broken for you."

Source: *Nicene and Post-Nicene Fathers*, 2nd series, Vol. 14, edited by Henry R. Percival. Peabody, MA: Hendrickson Publishers, 1885.

Meanwhile, the Antiochene party refused to have anything to do with the "Cyrillians," to which the main council under Cyril responded by excommunicating any bishop who "has joined himself to the assembly of revolt" (Stevenson and Frend 1989: 310).

Theodosius was still backing Nestorius at this point, but his support was wavering. On July 17, he ordered that Cyril and Memnon be deposed along with Nestorius. All three were arrested in August, after which both sides wrote letters to the emperor, appealing their cases. Finally, on September 11, 431, Theodosius dissolved the council, sending Nestorius back to his monastery in Antioch, while Cyril returned in victory to Alexandria, having overcome the imperial judgment against him by the distribution of extensive bribes to the court in Constantinople, a practice he repeated later on to maintain his position of favor with the imperial family and to ensure that the Antiochenes would agree to the *Formula of Reunion* in 433 (Bethune-Baker 1908: 10–1; Loofs 1914: 55–56; Driver and Hodgson 1925: 279– 82, 349–51).

Over the next two years, the emperor and representatives from the two sides in the conflict conducted negotiations aimed at reconciliation. Cyril was especially motivated to see this happen, for the Council of Ephesus could not be considered as binding unless there was unanimous agreement to its decision. Without the support of Antioch, the ecclesiastical legitimacy of Nestorius's deposition was questionable and Cyril's position was vulnerable. Finally, in April 433, in response to an Antiochene proposal, Cyril wrote a letter to John of Antioch "to make peace between the Churches" and agreed to the Formula of Reunion, probably drawn up by Theodoret of Cyrrhus.

The Formula was a compromise theological statement that favored the Antiochene "two-nature" position over the Alexandrian "one-nature" position and made no mention of Cyril's contentious anathemas, but confessed Mary as *Theotokos* (Stevenson and Frend 1989: 314–15). As Loofs points out, Cyril "could have come to an agreement with him [Nestorius] as easily as with the Antiochians afterwards in 433, if he had not had . . . an interest in discrediting him" (1914: 41). As it was, to secure the peace, the Antiochenes had to accept the decisions of Ephesus as binding, including the deposition of Nestorius. Thus, in exchange for a theological agreement that he would have whole-heartedly agreed with, Nestorius was sacrificed and thereafter considered a heretic.

That Cyril had not abandoned his essentially "one-nature" approach is evident from a letter he wrote to fellow Alexandrians to defend his acceptance of the Formula of Reunion, in which he stated unequivocally "after the union [the Incarnation] we do not divide the natures from each other . . . but say 'one Son' and, as the fathers have put it, 'one incarnate nature of the Word'." This final phrase was a quote from the heretical Apollinarius, which Cyril believed to be from Athanasius (Stevenson and Frend 1989: 318). Although the Antiochene bishops agreed to acknowledge Cyril as orthodox, many initially refused to accept the deposition of Nestorius. However, by 437, all had finally agreed to this, many albeit reluctantly. Nestorius had become expendable and denouncing him was the price of theological peace: "John of Antioch . . . and Pope Celestine of Rome ended up taking the side of Cyril against Nestorius, not for theological reasons, but for church-political reasons . . . there is no evidence that they held a different viewpoint from Nestorius. Actually, all the evidence indicates that they held precisely the same view" (Braaten 1963: 252).

Accusations against Nestorius and the *Bazaar of Heracleides*

The standard accusations against Nestorius can be summed up as follows:

1. By rejecting the term *Theotokos*, he ignored the importance of the *communicatio idiomatum* (the idea that all the attributes of divinity in Christ can be attributed to his humanity and vice versa) and challenged (or even denied) the divinity of Christ, presenting him rather as a "mere man."

2. By calling the union of divinity and humanity in Christ a "conjunction" of the two natures and promoting a "prosopic union" rather than Cyril's "hypostatic union," he devalued the idea that "the Word became flesh."

3. By differentiating between Christ's humanity and divinity, he promoted "two persons," "two Sons," and "two Christs," rather than a unified person.

Did Nestorius actually teach any of these things? In order to determine this, his extant writings need to be analyzed. Until the late 19th century, this task was particularly difficult, since only a few of his works remained, nearly all in carefully selected fragments preserved in the Acts of the Council of Ephesus or the writings of Cyril and others or disguised as sermons of John Chrysostom (Bethune-Baker 1908: 23–25; Nau 1910: 335–58; Driver and Hodgson 1925; 382–98). However, the discovery in 1889 of the *Bazaar of Heracleides*, a Syriac translation of Nestorius's defense of his life and doctrine, gave scholars new insights into the teachings of the condemned heretic. The work, originally composed in Greek under the pseudonym Heracleides (so as not to attract the attention of those intent on burning Nestorius's writings), was probably finished sometime between late 450 and late 451, around the time of the Council of Chalcedon (Driver and Hodgson 1925: x; Bevan 2007: 42). Although there are no references to Chalcedon in the book, it has been proposed that Nestorius was actually summoned to the council, but died en route (Bevan 2007: 42–51).

As Driver and Hodgson (the English translators of the work) note that Nestorius's aim was to show that "his own condemnation at Ephesus was unjust" and "the vindication of Flavian [after the 449 Council] . . . was the vindication of all that he [Nestorius] had stood for." Over and over again he makes the point that his doctrines are consistent with the Bible, the Nicene Creed, and the Church Fathers (Driver and Hodgson 1925: xxix–xxxi). Some have questioned how much the *Bazaar* accurately expresses his position 20 years earlier during the height of the controversy. However, as Anastos has noted, "it remains legitimate to allow him to be judged by his own latest and most mature efforts" (1962: 121).

All commentators on the *Bazaar* agree that there are significant problems in understanding the text. A major objection concerns the unity of the book, which can be divided into two parts: a *Dialogue* between Nestorius and Sophronius (Driver and Hodgson 1925: 7–86) and an *Apology* by Nestorius (Driver and Hodgson 1925: 87–380). Some have maintained that, whereas the *Apology* is unquestionably by Nestorius, the *Dialogue* is the work of a later author, pseudo-Nestorius (Turner 1975: 306–8), but this idea has been disputed by others (Chesnut 1978: 392–98). Most scholars note the serious stylistic challenges the work presents, challenges that in part explain why Nestorius's ideas were never broadly accepted: "It is not possible . . . to gather together a series of quotations from the Bazaar which, without explanation of linkage, will give

a coherent and connected account of the Incarnation" (Vine 1948: 188). Even Anastos, who regards him as "indubitably orthodox" and the "most brilliant theologian of the fifth century," describes the repetition in the *Bazaar* as "frustrating, wearisome, and painful" and concludes that his major defects were "the obscurity and prolixity of his style" (1962: 123, 140).

Nonetheless, careful consideration of both the *Bazaar* and the other extant fragments of Nestorius's writings can greatly help in dispelling some of the misunderstandings about his teaching that persist to this day. Although scholars continue to disagree over exactly what he taught, the work clearly shows that he denied (1) an essential union of the divine and human natures in Christ (i.e., a union of the essence or substance of each nature); (2) any transformation from Godhead to manhood or vice versa in the Incarnation; (3) the idea that Christ was just another "inspired man"; (4) the notion that either of the two natures in Christ was not real; (5) the suffering of the divine logos during the Incarnation; and (6) the idea of "two Sons" in Christ (Driver and Hodgson 1925: xxxii). Let us now analyze the key accusations against Nestorius with reference to his defense in the *Bazaar*.

Nestorius's Rejection of *Theotokos*

As noted above, Nestorius's rejection of *Theotokos* is seen by many as ignoring the importance of the *communicatio idiomatum* (the sharing of attributes between the divinity and humanity of Christ) and thus challenging the divinity of Christ. However, these accusations ignore several facts. First, Nestorius's objection to *Theotokos*, politically unwise as it was, was based on biblical statements that speak of Mary as the mother of "Jesus," "Christ," or "the Son of God," but not God. In contrast to those who viewed Mary "as in some kind of way divine, like God," Nestorius claimed to be following both "the holy fathers of Nicaea" and "the Scriptures" in his opposition to *Theotokos* (Bethune-Baker 1908: 17); indeed, he was on much more solid exegetical ground than Cyril and others who championed the term.

Second, although he preferred *Christotokos* to *Theotokos*, Nestorius did not completely exclude the use of the latter, as long as it was clarified. As he stated in a sermon: "If any of you or any one else be simple and has a preference for the term *Theotokos*, then I have nothing to say against it—only do not make a Goddess of the virgin" (Loofs 1914: 32; cf. Sellers 1940: 172–73). The subsequent growth of the cult of Mary in many parts of the Christian world, in which the "Mother of God" is referred to as the "Queen of Heaven" and treated virtually as a goddess, can be seen as a realization of Nestorius's fears. In particular, the later role of the *Theotokos* as "the special protectress of Constantinople" who "fought alongside them [the inhabitants] in the battle" during the Avar-Persian siege of the city in 626 (Cameron 1978: 78–79) would have made Nestorius turn in his grave.

Third, Nestorius's formula of "the divinity makes use of the prosopon of the humanity and the humanity of that of the divinity" essentially serves the same function as the *communicatio idiomatum* (Anastos 1962: 136; cf. Sellers 1940: 167–71), so that "the *Logos* shows himself in the form of a servant and the man in the form of God" (Loofs 1914: 83; cf. Bethune-Baker 1908: 95; Driver and Hodgson 1925: 190, 241). As Anastos notes, the difference between Cyril and Nestorius on this point concerned "their disagreement concerning the subject of the God-man's career and experience. Cyril . . . [following the Alexandrian Christology] preferred to begin with the divine Logos. . . . Nestorius . . . associates all these activities [suffering, dying, rising from the dead] with 'the prosopon of the union' (the Jesus Christ of the Gospels)" (Anastos 1962: 138; cf. Driver and Hodgson 1925: 141–48).

Finally, the equation of Nestorius's position on *Theotokos* with the heretical ideas of Paul of Samosata and Photinus (d. 376), both of whom denied the divinity of Christ, ignores that fact that Nestorius's motivation in opposing *Theotokos* was to protect the Godhead from being diminished, for "if the Godhead of the Son had its origin in the womb of the Virgin Mary, it was not Godhead as the Father's" and therefore was akin to Arianism (Bethune-Baker 1908: 19). In fact, Nestorius sought to avoid two erroneous ideas, that the Godhead had its origin in a human being (Mary) and that the manhood of Christ was somehow less real than that of humanity in general (Bethune-Baker 1908: 62).

A related charge that was made against Nestorius must also be mentioned here. He was accused in the Acts of the Council of Ephesus of having said "I could not give the name of God to one who was two or three months old," referring to the Christ child. This was accepted without further inquiry by Cyril as evidence of Nestorius's rejection of the divinity of Christ. However, based on Nestorius's own account in the *Bazaar*, it seems that his probable words were that "he could not bring himself to call God a babe. . . . He refused to predicate infancy of God, rather than Godhead of an infant," a crucial difference (Bethune-Baker 1908: 77; cf. Driver and Hodgson 1925: 136–41).

Nestorius's "Prosopic Union"

Although Nestorius's rejection of *Theotokos* was the flashpoint for the controversy (due largely to the popularity of the term), Cyril equally critiqued him over his notion of a prosopic union. Herein lies a key problem over which scholars continue to disagree: the nature of Nestorius's metaphysical system and its relation to the "orthodox" Chalcedonian view of two natures in one person (the latter represented by one *prosopon* and one *hypostasis* in the Chalcedonian definition). This is difficult to unravel, given the different ways that Nestorius, Cyril, and others in the fifth century used the relevant Greek terms. Following the differentiation between *hypostasis* and *ousia* introduced by Basil

of Caesarea 50 years earlier, Cyril located the "person" of Christ in the *hypostasis*. Thus for him, the union of divine and human was a "hypostatic union."

In contrast, Nestorius generally used *hypostasis* in the older sense, as a synonym for *ousia*. For him, both the divine and the human in Christ each had not only their own nature, but also their own *hypostasis*/*ousia*. Since the Nicene Creed had declared the Son to be *homoousios* (of one *ousia*) with the Father, Nestorius was unwilling to distinguish the *hypostasis* (equal to the *ousia* in his mind) of the Son from that of the Father, a necessary requirement for a hypostatic union to take place (since it only took place in the Son, not the Father or the Spirit). Because the *ousiai* of Godhead (which Christ shared with the Father and the Spirit) and manhood (which he shared with all humanity) were completely different essences, they could not be combined with each other; "To Nestorius Godhead and manhood . . . were much too real to be able to lose themselves in one another; the unity must be found in something other than the 'substances' themselves" (Bethune-Baker 1908: 53).

Rejecting Cyril's "hypostatic union" (which in Cyril's terminology also implied a union of natures), Nestorius opted instead for a "prosopic union," different in kind from both the unity of *ousia* (substance) shared by the members of the Trinity and the involuntary natural unity of body and soul in humans, which was used by Cyril as a metaphor for the divine–human union in Christ (Driver and Hodgson 1925: 412–13; Anastos 1962: 126–27). "Nestorius rejected the idea of a substantial union [because] such a union would result in a confusion of God and man" (Braaten 1963: 260) in which "each loses its own identity and ceases to function as a self-contained unit" (Chesnut 1978: 403). His starting place was quite different from Cyril's, as he notes in the *Bazaar*: "It is not the Logos who has become twofold; it is the one Lord Jesus Christ who is twofold in his natures. In him are seen all the characteristics of the God-Logos . . . and also all those of the manhood" (Loofs 1914: 79–80; cf. Driver and Hodgson 1925: 145).

Nestorius's theory of the prosopic union suggests that "in the person of Christ, a union of two persons took place so that they exchanged what is each other's . . . the union takes place in the interchange of roles, the one making use of the *prosopon* of the other" (Braaten 1963: 261). Thus, "the Logos 'takes' the prosopon of the manhood . . . as his prosopon, and 'gives' His divine prosopon to the manhood" (Sellers 1940: 147; cf. Driver and Hodgson 1925: 69–70). Or again, "Christ is the union of the eternal Logos and the Son of Mary, the principle of the union being that the *prosopon* of each has been taken by the other, so that there is one *prosopon* of the two in the union." In contrast, Nestorius terms Cyril's hypostatic union as "unscriptural, unorthodox, destructive of true religion, and unintelligible" (Driver and Hodgson 1925: xxxii–xxxiii), realizing that, if "the divine Logos . . . took in his *hypostasis* a human body, soul and intellect . . . so that his human nature had, therefore, no *hypostasis*," the practical result was "a suppression of the manhood of Christ" (Loofs 1914: 72–73).

Although his critics, including Cyril, have typically rejected Nestorius's use of "conjunction" (Greek *synapheia*) as too weak to describe the relationship between the divine and human in Christ, Bethune-Baker notes that the word can also have the stronger meaning "contact" or "cohesion" and that Nestorius uses "united" and "union" more frequently in the *Bazaar* than "conjoined" or "conjunction." Throughout, Nestorius's main concern was to avoid "words like 'mixture', 'commingling', 'blending together', 'confusion' and . . . all ideas which would merge the two substances and natures of Godhead and manhood in one" (Bethune-Baker 1908: 91), resulting in either "an Arian doctrine . . . which makes of the Logos a creature [or] an Apollinarian doctrine . . . which renders the humanity incomplete" (Braaten 1963: 260). Thus, for Nestorius, "God the Word does not become in his very nature something that he was not before . . . [and] the man remains genuinely man within the incarnation" (Chesnut 1978: 407).

But what exactly did Nestorius mean by *prosopon*? Biblical and patristic writers before him had used it to convey the whole range of meaning noted above (face, mask, role, outward appearance, person), and three (Athanasius, Epiphanius, and Theodore of Mopsuestia) had used it to describe the Incarnation in ways that anticipated Nestorius's later use of the term (Driver and Hodgson 1925: 402–10). However, given this range of meaning and the fact that none of the terms Nestorius uses are exactly equivalent to our word "person" and the modern psychological framework it represents (Driver and Hodgson 1925: 412), it is misleading to automatically interpret *prosopon* in his works as "person." By doing so, we are in danger of evaluating him from our modern point of view, modified by nearly 1600 years of theological, philosophical, and psychological development since his time (Braaten 1963: 261).

Loofs suggests rather that "the main thing in his notion of *prosopon* . . . was the external undivided appearance" and specifically "the undivided appearance of the historic Jesus Christ" (1914: 76, 79), an idea expanded by Driver and Hodgson, who propose that "Nestorius analysed everything that exists into . . . essence [*ousia*], nature [*physis*] . . . and appearance [*prosopon*]," the latter being "a real element in the being of a thing." As such, the prosopic union was not merely a "moral union" but a "real metaphysical unity," although Driver and Hodgson suggest it was "not strong enough to bear the strain it was designed to meet," to explain the oneness of Godhead and manhood in Christ (1925: 414–17, 419).

Chesnut further observes that "to be the prosopon of God means to Nestorius to be the Image of God, and to be the Image of God is first and foremost to will what God wills, to have the will and purpose of God" (1978: 399; cf. Driver and Hodgson 1925: 59; Sellers 1940: 134). This aspect of *prosopon* reminds us of the Antiochene emphasis on the union being voluntary, requiring the active participation of Christ's human nature. As Turner notes, "the problem is vital for Nestorius but purely marginal for Cyril" (1975: 311). However, the presence of a human will in Christ does not jeopardize the will of God, for as Nestorius explains, "he

[Christ] acquired nothing else than to wish and to will whatever God willed in him. For this reason God became in him whatever he was in himself" (Chesnut 1978: 400; cf. Driver and Hodgson 1925: 251). Again, "he in nothing deviated from the purpose of God . . . his will was bound to the will of God" (Chesnut 1978: 401; cf. Driver and Hodgson 1925: 63–64; Sellers 1940: 138–40).

Nestorius and "Two Persons" in Christ

The contrast between Alexandrian and Antiochene thought outlined above was essentially a difference in emphasis, between the unity of the person of Christ (Alexandria) and the duality of his divine and human natures (Antioch). Both schools of thought tended to accuse the other of overstating their respective emphasis. That Nestorius and other Antiochenes were accused of preaching "two persons" in Christ is therefore not surprising; this misinterpretation typically occurs when "the context and characteristics of the Christological language of the Antiochene tradition are ignored" (Uthemann 2007: 477). One of the main problems seems to have been Alexandria's inability to accept the symmetrical Christology of Antioch, where divinity and humanity both played key roles, united in the person of Christ. By contrast, Cyril and other Alexandrians insisted on the subject of their Christology being the divine logos, with the result that Christ's humanity became less important. Any attempt by Nestorius or other Antiochenes to present a balanced picture was interpreted as "preaching two persons."

However, Nestorius expressly denies any belief in two sons or two christs, ascribing this view to the followers of Paul of Samosata ("They speak of a double son and a double Christ"). In an exposition of the introduction to John's gospel, which refers to the divine Word of God indwelling Christ, he says, "How then can we understand this to be one Son, and Christ to be another Son, and one that is man only?" Elsewhere, he remarks, "God the Word and the man in whom He came to be are not numerically two" and "He is a single (person), but . . . He is different in the natures of manhood and Godhead" and "I call Christ perfect God and perfect man, not natures which are commingled, but which are united" (Bethune-Baker 1908: 82–85; cf. Driver and Hodgson 1925: 45–46, 50).

Thus, judged by his own words, Nestorius comes across not as a heretic, but as orthodox, in agreement with the theology articulated at Chalcedon. Indeed, he was in complete accord with the Tome of Leo, commenting when he read it, "I gave thanks to God that the Church of Rome was rightly and blamelessly making confessions, even though they happened to be against me personally" (Bethune-Baker 1908: 191–92; cf. Driver and Hodgson 1925: 340). A letter of Nestorius to the inhabitants of Constantinople, probably from 449, further states: "It is my doctrine which Leo and Flavian are upholding. . . . Believe as our holy comrades in the faith, Leo and Flavian!" (Loofs 1914: 25).

Nonetheless, Nestorius's use of *prosopon* is sometimes confusing and undoubtedly supported his enemies' accusations. Besides describing the union occurring in one *prosopon*, he also refers in places to two *prosopa* in Christ, although the former use is much more common than the latter (Loofs 1914: 79). Anastos concludes that he used *prosopon* in two distinct senses: (A) "the exterior aspect or appearance of a thing" (as Loofs observed) and (B) "an approximate equivalent of our word 'person'." The first relates to the two natures of Christ, indicating "each had a substantive reality . . . which remained undiminished after the union," while the second relates to Jesus Christ as "the common prosopon of the two natures." Nestorius is then able to speak of the "two *prosopa (sense* A) . . . in the one *prosopon (sense* B) of Jesus Christ" (Anastos 1962: 129–30; cf. Chesnut 1978: 402; Uthemann 2007: 478).

Put another way, "Nestorius' theory was that the two distinctly existing persons combine to make a new person, who is called Jesus. Hence, Jesus is one person made up of two persons" (Braaten 1963: 258). Admittedly, this dual sense of the word, never clearly explained by Nestorius, is confusing and opens him up to criticism, but given the general fluidity in the terminology of "personhood" mentioned above, it is not surprising and should not be grounds for accusing Nestorius of heresy, especially when he openly said "I separate the natures; but unite the worship" (Sellers 1940: 196).

Based on this distinction in the use of *prosopon*, Anastos summarizes Nestorius's actual Christology as follows:

> Jesus Christ was the divine Logos incarnate, the Son of God in the flesh, the Lord whom his disciples knew as a man but recognized to be God. The unity of his "personality" was further guaranteed by the fact that it was the Logos who both "gave" his prosopon (sense A) to the human nature and "took" that of the human for his own. Moreover, the human will of Christ was always obedient to the divine, so that there never was any conflict or division between the two. (Anastos 1962: 132)

Anastos further comments that "Nestorius' Christology is not characterized by preoccupation with either one of the two natures to the exclusion or detriment of the other, but rather by uncompromising insistence upon the union of both of them in Christ, in their full totality, and unimpaired" (1962: 140).

The Aftermath

In 435, Theodosius had officially banned "the impious books of the said lawless and blasphemous Nestorius" and had forbidden his followers "all right of assembly," an edict that was reissued in modified form in 448, during the height of Theodosius's subsequent support for Eutyches (Millar 2006: 176–77, 186–87). Nestorius himself was banished to Arabia in 436, eventually ending

up in the Egyptian desert. The deaths of John of Antioch (440) and Cyril (444) ended this chapter in church history. Although the Council of Chalcedon in 451 virtually eulogized "the blessed Cyril" in its Definition document, it anathematized "those who feign that the Lord had two natures before the union, but . . . one after the union," a perfect description of Cyril's position (Stevenson and Frend 1989: 352). Nonetheless, for the sake of ecclesiastical unity, the rallying cry was "Cyril and Leo taught alike" (Frend 1972: 48), and Nestorius continued to be the scapegoat, even though the language of the Definition of Faith, "apart from the word '*hypostasis*'. . . was exactly that used . . . by the West and by Nestorius" (Gray 2005: 222). Indeed, paradoxically, "the essence of Nestorius' beliefs, without his name attached to them, came to be affirmed at the Council of Chalcedon under a Cyrillian guise" (Bevan 2007: 40). Again, although the Western position coincided theologically more with the Antiochene position, "tactically and emotionally Rome was the ally of Alexandria" (Frend 1972: 131–34), and in the end, tactics trumped theological consistency.

By this time, Cyril's status as the champion of orthodoxy had become virtually unassailable; "at Chalcedon and for the century after each party [Chalcedonian or Miaphysite] was able to claim Cyril for their own and set one quotation from his works against another" (Frend 1972: 23). This universally favorable view of Cyril after Ephesus has traditionally been interpreted as evidence of the superiority of his theological views over those of Nestorius, but some have suggested that it equally reflects his polished rhetorical skills: "Nestorius' homiletic discourse was pedantic and recondite in style, while Cyril's was lively" (Wessel 2004: 9). Although some in Constantinople were concerned over Cyril's references to "one incarnate nature" in Christ, especially in light of his equation of *hypostasis* with *physis* (so that his "one *hypostasis* in Christ" could easily be interpreted as "one nature in Christ"), these objections were overcome by "Cyril's mastery of rhetorical argumentation" (Wessel 2004: 298, 301). Throughout this time, up to his death sometime after 450, Nestorius remained in exile in Egypt, well informed of ecclesiastical developments, as we learn from his extant memoirs, to which we now turn.

Conclusion

In conclusion, a comment is necessary about the scholarly approach to Nestorius, especially in the 20th century. Several scholars have concluded that Nestorius was either entirely or nearly orthodox in his beliefs, including Bethune-Baker (1908), Loofs (1914), Sellers (1940), Vine (1948), and Anastos (1962). Indeed, after reading the *Bazaar*, it is clear that he was not guilty of the heresy he was accused of, namely preaching two persons in Christ. Again, note the almost Chalcedonian ring of his confession of "one Christ, one Son, one Lord," and "in one Christ two natures without confusion. By one nature . . .

of the divinity, he was born of God the Father; by the other . . . of the humanity, [he was born] of the holy virgin" (Anastos 1962: 128; cf. Driver and Hodgson 1925: 295–96). Both Loofs and Anastos conclude that, being in full agreement with the Tome of Leo, Nestorius would have wholeheartedly approved of the Chalcedonian confession (Loofs 1914: 99–100; Anastos 1962: 138; cf. Driver and Hodgson 1925: 388–89).

However, the a priori conviction of other scholars that Nestorius was a heretic no matter what he actually taught has made it impossible for them to revise their views: "The old notion that church councils cannot err seems to exercise a powerful influence on some scholars" (Braaten 1963: 254). For those who believe that all church councils have the same divine inspiration and authority as the first Council of Jerusalem described in the New Testament (Acts 15:28), the idea that the church fathers may have gotten it wrong is a threatening concept.

Such scholars dismiss statements by Nestorius that are in fact orthodox as "an emergency invention forced upon him by his adversaries" or as evidence that he "used orthodox phraseology to confound his readers, or he used the orthodox terms in an ambiguous sense, meaning something else by them" (Braaten 1963: 255). They tend to judge Nestorius "in terms of "orthodox christological categories which were made precise at a later date," one even suggesting that if the orthodox Catholic position on Nestorius is questioned, then "even the doctrine of the infallibility of the Pope is at stake" (Braaten 1963: 260). Such a subjective approach is intellectually dishonest and patently unfair to Nestorius.

In contrast, Nestorius's words in the *Bazaar* provide a fitting conclusion to this tragic chapter in church history: "The goal of my earnest wish, then, is that God may be blessed on earth as in heaven. But as for Nestorius, let him be anathema. . . . And would to God that all men by anathematizing me might attain to a reconciliation with God; for to me there is nothing greater or more precious than this" (Bethune-Baker 1908: 190, 198; cf. Driver and Hodgson 1925: 372).

References and Further Reading

Anastos, Milton V. "Nestorius Was Orthodox." *Dumbarton Oaks Papers* 16 (1962): 117–40.

Bethune-Baker, J. F. *Nestorius and His Teaching: A Fresh Examination of the Evidence*. Cambridge: Cambridge University Press, 1908.

Bevan, George A. "The Last Days of Nestorius in the Syriac Sources." *Journal of the Canadian Society for Syriac Studies* 7 (2007): 39–54.

Braaten, Carl E. "Modern Interpretations of Nestorius." *Church History* 32 (1963): 251–67.

Brock, Sebastian P. "The 'Nestorian' Church: A Lamentable Misnomer." *Bulletin of the John Rylands University Library* 78, 3 (1996): 23–35.

Cameron, Averil. "The Theotokos in Sixth-Century Constantinople." *Journal of Theological Studies* 29 (N.S.) (1978): 79–108.

Chesnut, Roberta C. "The Two Prosopa in Nestorius' *Bazaar of Heracleides*." *Journal of Theological Studies* 29 (N.S.) (1978): 392–409.

DelCogliano, Mark, trans. "Nestorius, 2nd and 3rd Letters to Pope Celestine." www.tertullian.org/fathers/nestorius_two_letters_01.htm (translated 2005, cited April 5, 2009) (accessed May 31, 1010).

Driver, G. R., and Leonard Hodgson, trans. *Nestorius: The Bazaar of Heracleides*. Oxford: Clarendon, 1925.

Frend, W. H. C. *The Rise of the Monophysite Movement: Chapters in the History of the Church in the Fifth and Sixth Centuries*. Cambridge: Cambridge University Press, 1972.

Gray, Patrick T. R. "The Legacy of Chalcedon: Christological Problems and Their Significance." In *The Cambridge Companion to the Age of Justinian* (pp. 215–38). Edited by Michael Maas. Cambridge: Cambridge University Press, 2005.

Loofs, Friedrich. *Nestorius and His Place in the History of Christian Doctrine*. Cambridge: Cambridge University Press, 1914. (Reprint: New York: Burt Franklin Reprints, 1975).

Loofs, Friedrich, ed. *Nestoriana: Die Fragmente Des Nestorius*. Halle: Max Niemeyer, 1905.

McGuckin, J. A. "Nestorius and the Political Factions of Fifth-Century Byzantium: Factors in His Personal Downfall." *Bulletin of the John Rylands University Library* 78, 3 (1996): 7–21.

Millar, Fergus. *A Greek Roman Empire: Power and Belief under Theodosius II (408–450)*. Berkeley: University of California Press, 2006.

Nau, François, trans. *Nestorius: Le Livre d'Héraclide de Damas*. Paris: Letouzey et Ané, 1910.

Russell, Norman. *Cyril of Alexandria*. London: Routledge, 2000.

Sellers, R. V. *Two Ancient Christologies*. London: SPCK, 1940.

Stevenson, J., and W. H. C. Frend, eds. *A New Eusebius: Documents Illustrating the History of the Church to AD 337*. 2nd ed. London: SPCK, 1987.

Stevenson, J., and W. H. C. Frend, eds. *Creeds, Councils and Controversies: Documents Illustrating the History of the Church AD 337–461*. 2nd ed. London: SPCK, 1989.

Turcescu, Lucian. "Prosōpon and Hypostasis in Basil of Caesarea's 'Against Eunomius' and the Epistles." *Vigiliae Christianae* 51, 4 (1997): 374–95.

Turner, H. E. W. "Nestorius Reconsidered." In *Studia Patristica XIII* (Texte und Untersuchungen 116) (pp. 306–21). Berlin: Akademie-Verlag, 1975.

Uthemann, Karl Heinz. "History of Christology to the Seventh Century." In *The Cambridge History of Christianity*, Vol. 2: *Constantine to c. 600* (pp. 460–500). Edited by Augustine Casiday and Frederick W. Norris. Cambridge: Cambridge University Press, 2007.

Vine, Aubrey R. *An Approach to Christology*. London: Independent Press, 1948.

Wessel, Susan. *Cyril of Alexandria and the Nestorian Controversy*. Oxford: Oxford University Press, 2004.

CON

There were dramatic changes in the Roman world during the fourth and fifth centuries. In the western empire, invasions from cultures as diverse as the Visigoths, Vandals, and Huns marked this time period as extremely unstable. The eastern empire, however, weathered the storm slightly less dramatically. Because of the strategic location of the eastern capital of Constantinople, a wealthier agricultural base, and fewer vulnerable frontiers, the eastern empire was able to survive for another thousand years. Yet even the eastern empire experienced episodes of instability, and these periods of unrest are mirrored in church history.

The church of the fifth century was organized into a patriarchate model, with three bishops serving as patriarchs of the church. These patriarchs were more powerful than other bishops in the church, and they included the bishop of Rome, the bishop of Alexandria, and the bishop of Antioch. The bishop of Constantinople and the bishop of Jerusalem were in the second tier and were elevated to patriarch status by the Council of Chalcedon in 451, leading to something called the pentarchy. Eventually, the capital cities of Rome and Constantinople featured the two principal bishops.

Throughout the fifth century there were arguments about the definition of acceptable and unacceptable belief systems for the developing Christian religion. One of a series of recurring questions concerning the Christology of Christ was: Was Christ human, divine, or both? There were several church councils called to debate this important issue, including the Council of Nicaea (325) among others. One of the principal participants in the controversy was Nestorius (ca. 381–ca. 451), the bishop of Constantinople. We know little about Nestorius's early life, but according to the theological historian Friedrich Loofs, Nestorius was born in Syria and became a monk at an early age. He studied under Theodore of Mopsuestia (ca. 350–428) for a time. Nestorius was living as a monk when he was chosen by the weak leader Emperor Theodosius II (Flavius Theodosius II [401–450]) to become the bishop of Constantinople in 428 (Loofs 1914: 7).

Nestorius was a controversial figure from a very early age. As a young man, he was well known for his ascetic practices and for his public speaking

ability. As a result of a power struggle, he was selected to be the bishop of Constantinople by Theodosius II. Even after his promotion to bishop, Nestorius maintained strong ties to Antioch. Immediately upon assuming his new leadership position, Nestorius began a series of reforms to make certain parishioners in Constantinople practiced Christianity correctly. As one of his first acts as bishop, he expelled several groups from Constantinople, including Arians and Novatianists. Then, he took disciplinary actions against a number of monks, including those who were not associated with particular monasteries, wandering monks, and monks who were acting incorrectly (Caner 2002: 212–15). Next, in one of his first sermons, Nestorius preached against the popular practice of referring to the Virgin Mary as *Theotokos* (which means "God-bearer" or "Mother of God"). According to Nestorius, the problem with calling the Virgin Mary the "God-bearer" was that Mary was human and therefore could not have been the mother of the divine nature of Christ (Nystrom 2003: 94). He worried that by using the term *Theotokos* the full humanity of Christ might be compromised. All of these actions served to bring him to the attention of rival bishops in the area.

The controversy concerning the correct title for the Virgin Mary actually had its beginnings in Nestorius's hometown—the city of Antioch. According to historian Philip Rousseau, Antioch was the one city that represented the new Christian world and was quite famous as the first place that followers of Christ were called "Christians." Through the years, Antioch housed a number of famous theologians including John Chrysostom (Rousseau 2002: 349–407) and Nestorius. Antioch was famous for the number of theological disputes that arose there as well. Additionally, Antioch was an important region politically, acting as a border between the eastern empire and the Sassanian Persians. Because of its historical and political position, Antioch became a central clearinghouse for new ideas concerning the changing views of Christianity. Unfortunately, failure to resolve the controversies concerning these shifting ideas eventually drove a wedge between Antioch and other Christians in the east (Rousseau 2002: 191–92).
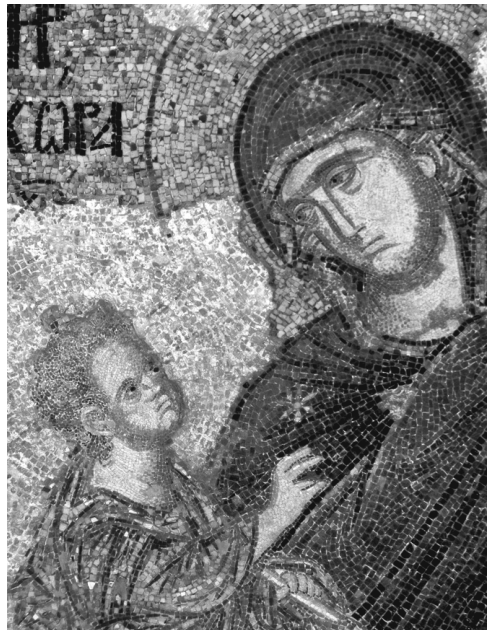
Mosaic of Virgin Mary and Jesus Christ, Hagia Sofia, Istanbul. (Pavle Marjanovic/Dreamstime.com)

Historical Background

Most of the conflicts concerning Christianity within the city of Antioch resulted from a dispute that had arisen over 100 years earlier in nearby Alexandria concerning the divinity of Christ. This argument was known as the Arian controversy, which developed into one of the most long lasting of all the Christian heresies. We know little about the founder of this debate, because it is difficult to reconstruct the life and writings of many early Christian writers, since they were later deemed heretical and their writings destroyed. Most of the information on Arius can be found in the writings of Athanasius (ca. 293–373) and the church historians Rufinus, Socrates, Sozemen, and Philostorgius. Beginning around 318, a priest named Arius clashed with Alexander, the bishop of Alexandria, over the interpretation of the divinity of Christ. This debate focused on the identity of Christ, centering on the first chapter in the Gospel of John. Arius claimed that the Son of God was less divine than God but was created by God before time began and was therefore a creature of God. For Arius, the Son was not made of the same divine "essence" of the Father and was less divine than God, but the Son still was divine. This argument was problematic because it seemed to indicate the existence of two gods. Arius also argued that the Son had a diminished role and that God had produced the Son "in time"; therefore, the Son was noneternal, and susceptible to change. If the Son was created "in time," then, at least in theory, the Son would be capable of change because nothing created in time can be expected to last indefinitely. The prospect of a changing Son was extremely troubling because it seemed to threaten the concept of eternal salvation (Nystrom 2003: 90).

Arius's teachings spread rapidly throughout the region; assisted no doubt by an Arian saying that appeared in popular culture as a song. By the time Bishop Alexander condemned his teachings and excommunicated him (ca. 320), Arius had gone into exile. While exiled, Arius's influence grew and he won new adherents, including several powerful eastern bishops. Despite efforts by various bishops to quell its popularity, the Arian movement continued to spread and eventually became powerful enough to threaten a permanent division between Christian factions. In order to maintain peace in the recently fused empire, the Emperor Constantine called a General Council at Nicaea in 325. It was here that the opponents of Arianism, led by the deacon Athanasius, succeeded in defining the coeternity and coequality of the Father and the Son, using the term *homoousios* to describe their sameness of substance. They also decided that the Son was not created, but instead was begotten, and that the Son had always existed. These beliefs were woven into the Creed of Nicaea and were explicitly included as a reaction to Arian beliefs. At the conclusion of the Council at Nicaea, Arius and several of the bishops who supported him were banished.

Unfortunately, the Council and Creed of Nicaea did not put an end to the controversy surrounding the relationship of the Father and the Son. After the

death of Constantine in 337, the disagreements resumed. The Arian movement was still strong enough to sway the opinion of the general population, some of the eastern bishops, and even the new emperor, Constantius II (337–361). These bishops made the argument that the term *homoousios* did not allow enough distinction between the Father and the Son, because of its emphasis on the Father and the Son's sameness of substance. The bishops argued that *homoiousios*, "of similar substance," was a better definition. The eastern bishops lined up against the Nicene bishops, led by Athanasius until he died in 373. After his death, three great theologians from Cappadocia (northern Turkey) continued to speak for the decisions that were reached at Nicaea. Basil (the Great), Gregory of Nyssa, and Gregory of Nazianzus became known as the Cappadocian Fathers. They wrote their opinions in Greek and argued that it is the nature of divinity to express itself as a triune entity, and that the Father, Son, and Holy Spirit were equal in divinity, yet at the same time were distinctly individual. This equality eventually became the idea of the Trinity, which was accepted as orthodox at the Council of Constantinople in 381. Because of this council, Arianism was declared a heresy and eventually died out.

After defining the relationship between the Father and the Son, the Christian theologians tackled issues relating to the nature of Christ and his role in salvation. This question was no less problematic than that of defining the Trinity. There were two schools of thought regarding the idea of the nature of Christ. The Antioch theologians suggested that Christ had two completely separate natures, both human and divine. The Alexandrian theologians took a different approach, choosing to emphasize only the divine nature of Christ. In order to solve the dilemma of which school of thought was correct, the Council of Chalcedon was called in 451. Unfortunately, this council did not fully solve the dilemma of the nature of Christ, and because they were unable to come to a conclusion, a Syrian bishop by the name of Apollinarius (ca. 310–ca. 390) began to teach that the Son's body was human but his mind was divine. For Apollinarius, this solution made perfect sense, as it demonstrated that humanity and divinity could be present in the same body.

Of course there were many reactions to the Apollinarius theory. The Cappadocian Fathers said that Christ had to become human so that individuals might be saved. Because Apollinarius said that Christ's mind and body were not of the same nature, it implied to the Cappadocians that Christ had not become completely human. Theodore of Mopsuestia (ca. 350–428) was also a critic of Apollinarius. He argued that Christ must have had a completely human mind because of his experiences of emotion and intellectual growth. Apollinarius was discredited in the Council of Constantinople in 381.

Theodore of Mopsuestia was from Antioch and had his own theory about the nature of Christ. He posited that Christ was the only person who had two complete natures. Each of the natures was clearly distinguishable from the

other, and this could be demonstrated from passages in the New Testament scriptures. According to Theodore, Christ's human nature was present when he wept or hungered—clearly physical manifestations of his human nature—and his divine nature was demonstrated when he performed miracles or taught parables. The focused emphasis on Christ's human nature allowed theologians to highlight the reality of Christ's human suffering, an emphasis that the Antioch school believed was often minimized by the Alexandrian school.

As discussed above, Nestorius was a student of Theodore of Mopsuestia, and because of this close association he carried Theodore's logic to the next step. According to scholar John McGuckin, Nestorius believed that the divine and the human parts of Christ were represented by two distinct natures—the divine nature demonstrating itself in divine works and the human nature as revealed by his earthly limitations (McGuckin 1994: 134–36). There was a popular custom of calling the Virgin Mary the "God-bearer" or "Mother of God," and Nestorius argued the imprecision of this designation. Since Christ experienced a human birth, this part of his nature should be categorized with physical characteristics like weeping or hungering. And since Mary could not be the mother of the divine nature in Christ—that part certainly came from the Father—Nestorius felt that it was not correct to call her *Theotokos*. A better title for the Virgin Mary, according to Nestorius, was *Christotokos* or "Christ-bearer/ Mother of Christ." This title recognized Mary as the mother of the human, though not of the divine nature of Christ.

Of course there was much controversy surrounding Nestorius's approach. Many Christians felt that Nestorius was attacking the cult of the Virgin Mother, and the acceptance of the title *Theotokos* became a sort of litmus test for faith. By 429 Nestorius had decided to bring his argument to the people by giving a series of public lectures at the cathedral. The first one of these sermons was given by a chaplain named Anastasius. In order to make sure Mary was not viewed as a sort of goddess figure, his lecture portrayed Mary as fully human, and he noted that it was impossible for God to be born from a mere woman. This sermon received very negative reviews, and another clergyman, Bishop Proclus (d. ca. 447) answered it with a sermon of his own on the Virgin Mother of God, which reiterated the popular view of the Virgin Mary. Since Proclus's sermon was greeted with loud applause, Nestorius decided to take matters into his own hands and answered the criticism with a series of his own sermons. The church historian Socrates suggests that by this time Nestorius saw this discussion as an underlying rebellion, symptomatic of those in Constantinople who resisted his reforms.

Nestorius's series of sermons at the cathedral did nothing to improve his standing in the theological community. Because of his stance on the *Theotokos* title, his opponents compared him to an ancient heretic Paul of Samosata (200–275), who taught that Jesus was only a God-inspired man. Rumors flew, claiming that Nestorius was teaching a theory of "two sons" rather than the orthodox

belief in one son, and bishops from surrounding areas began to take notice of the tumult.

The issue of whether or not to grant Mary the title of *Theotokos* continued as part of the controversy; but another issue, concerning the nature of the union between God and man in Christ, soon became more important. Nestorius taught that there were three *prosopa* (persons) living in Christ. These persons included a divine *prosopon*, a *prosopon* of the human nature or substance, and a union *prosopon*. For Nestorius, if Jesus was wholly God and completely human, there must be a *prosopon* of each nature and also a *prosopon* of the union. This third type of *prosopon* was theoretically defined by Nestorius as a sort of "conjunction" to describe the union of God and humanity (Urban 1986: 85–86). It was this third idea of the "union" person living in Christ that would prove to be Nestorius's undoing.

Historical Evidence

It is impossible to trace the life and ideas of Nestorius without also examining the life of Cyril of Alexandria (378–428). Cyril was one of the finest Christian theologians of his day, and he also stands out in the ranks of the greatest patristic writers (McGuckin 1994: 1). We know there were difficulties between Cyril and Nestorius because we have fragments of Nestorius's writings and sermons as well as Cyril's correspondence from the surviving Acts of the Council of Ephesus. Cyril decided it was time to intervene when Nestorius began to make imprecise statements. Prior to the Council of Ephesus in 431, Nestorius had announced to a group of bishops that God was not an infant two or three months old. What he had meant to say was that human suffering should be attributed to Christ's human nature, not his divine nature. Unfortunately, the bishops believed Nestorius meant that God could not have appeared in human form. Cyril did not hesitate to take full advantage of this and arranged for a public sign to be carried around town printed with Nestorius's words "Jesus was not God," openly accusing Nestorius of heresy (Wessel 2004: 140–42). In addition to the placard, several ascetic monks wrote a series of letters to the emperor asking for a trial for Nestorius on the charge of heresy and negligence toward church affairs. Since several of these monks were part of his own church, Cyril considered it his right to get involved.

Cyril was certainly brilliant, ruthless, and quite politically savvy. Even though Nestorius held the powerful office of bishop and had a reputation as a gifted speaker, he could not compete with Cyril. Beginning with a letter to monks in the area, Cyril began his crusade against Nestorius. This letter, in turn, elicited a response from one of his Nestorius's colleagues in order to answer to the points raised by Cyril. The correspondence continued between Nestorius and Cyril over the next two-and-a-half years.

In order to decide the matter, both Cyril and Nestorius appealed to Pope Celestine I (422–432). In August 430, Celestine called a church council in Rome and heard the case. The council ruled that *Theotokos* was the correct designation for the Virgin Mary, and that Nestorius should renounce his former views on the subject. Nestorius and his allies rejected the ruling and asked that the emperor Theodosius organize a general church council.

By the end of the summer of 431, Theodosius convened the Council of Ephesus to resolve the problem. Theodosius believed that Ephesus was a neutral site, and expressly desired Nestorius to be present at the meeting. Unfortunately for Nestorius, Cyril began the proceedings without waiting for either Nestorius or the bishops from Antioch to appear. When Nestorius and his allies arrived, they were met by a mob of irate monks organized by Cyril. Because of Cyril's skillful leadership, it only took one day to condemn Nestorius as the new "Judas," strip him of his authority as bishop, and defend the concept of the *Theotokos* as Mother of God. Nestorius was ostracized and banished into Egypt. He died in 451, always claiming that he had been misunderstood and that his downfall had been orchestrated by Cyril.

Religious fervor in Constantinople calmed down considerably after Nestorius's dismissal in 431, but divisions remained between Antioch and Alexandria. The Antioch school of theology continued to defend Christ's distinct humanity, and the Alexandrian school continued to teach that there was no distinction between the human and divine natures of Christ. Pressure was placed on Cyril to moderate his position in order to restore good relations with the church at Antioch. Accordingly, Bishop John of Antioch (428–441) wrote a compromise known as the *Formula of Reunion* (433), which described Christ as a union of both natures: divine and human. Cyril was later able to describe Christ as having one nature (hypostasis) arising out of two distinct natures. These ideas became known as the *communication idiomatum*—a sharing or imparting of particular qualities—and helped explain that Mary had given birth to a human being and had borne divinity as well.

The *Formula of Reunion* compromise remained in effect while Cyril and John of Alexandria were alive, but after they died conflict broke out again between Alexandria and Antioch. This time the disagreement concerned a theory by a monk from Constantinople named Eutyches. He maintained that Christ had two natures before the incarnation but claimed that they were so well blended in their union that afterward Christ only had a monophysite (one) nature. By this time, Dioscorus had taken over Cyril's post as bishop of Alexandria. He called another Council at Ephesus in 449 that verified Eutyches's theory of Christ's single nature and prohibited anyone from mentioning the concept of Christ's two natures. In order to back up his prohibition, Dioscorus brought a group of fanatical monks along to intimidate dissenters. This second Council at Ephesus was considered a travesty and was nicknamed the "Robber Council" by Pope Leo I (440–461) and appealed mainly to monophysite

---

The Canons of the Council of Ephesus

At the first Council of Ephesus in 431, Cyril of Alexandria was able to push through a series of holy laws, or canons, to ensure that Nestorianism would be considered a heresy, and all who subscribed to it heretics.

Canon II. IF any provincial bishops were not present at the holy Synod and have joined or attempted to join the apostasy; or if, after subscribing the deposition of Nestorius, they went back into the assembly of apostates; these men, according to the decree of the holy Synod, are to be deposed from the priesthood and degraded from their rank.

Canon III. IF any of the city or country clergy have been inhibited by Nestorius or his followers from the exercise of the priesthood, on account of their orthodoxy, we have declared it just that these should be restored to their proper rank. And in general we forbid all the clergy who adhere to the Orthodox and Ecumenical Synod in any way to submit to the bishops who have already apostatized or shall hereafter apostatize.

Canon IV. IF any of the clergy should fall away, and publicly or privately presume to maintain the doctrines of Nestorius or Celestius, it is declared just by the holy Synod that these also should be deposed.

Canon V. IF any have been condemned for evil practices by the holy Synod, or by their own bishops; and if, with his usual lack of discrimination, Nestorius (or his followers) has attempted, or shall hereafter attempt, uncanonically to restore such persons to communion and to their former rank, we have declared that they shall not be profited thereby, but shall remain deposed nevertheless.

Canon VI. LIKEWISE, if any should in any way attempt to set aside the orders in each case made by the holy Synod at Ephesus, the holy Synod decrees that, if they be bishops or clergymen, they shall absolutely forfeit their office; and, if laymen, that they shall be excommunicated.

Source: *Nicene and Post-Nicene Fathers*, 2nd series, Vol. 14, edited by Henry R. Percival. Peabody, MA: Hendrickson Publishers, 1885.

Christians. To answer the monophysites, another council was called at Chalcedon in 451. This council adopted a definition of Christology positing that Christ was one person possessing two natures, both human and divine, and each shared characteristics of the other. They also put forward that Christ was of one substance with the Father, thus silencing the Arians. The Council of Chalcedon was considered a success because it defined Christology in a way that most Christians consider orthodox even today. Those churches that were unhappy with the compromise position joined the monophysite churches, a tradition that remains today in the Coptic Christian communities of Egypt and Ethiopia. Other followers of Nestorian founded a group called Nestorianism. It managed

to continue by expanding eastward into Mesopotamia, Persia, and China. Nestorian communities can still be found today in Syria, Iraq, and Iran (Nystrom 2003: 94–96).

Historical Interpretation

Scholars who have analyzed the Nestorian controversy during the fifth century have questions about what the central question really is. As early as the fifth century an ecclesiastical historian named Socrates, in his *Ecclesiastical History*, viewed Nestorius as a victim of political pragmatism. He explained that even though Nestorius might be guilty of pride and ignorance, he was not guilty of heresy. Socrates restated Nestorius's view that the term *Theotokos* was problematic, as it might give rise to a goddess/virgin mythology. He also reiterated that Nestorius did not deny Christ's divinity. Martin Luther also examined the writings of Nestorius. In his book *Von Conciliis und Kirchen*, Luther notes that he did not understand the nature of Nestorius's error. After doing his own research on the matter, Luther found that Nestorius had been wrongly accused of teaching that Christ had two persons (Braaten 1963: 252–53).

Modern Historiography

There have been a number of texts published in the 19th and 20th centuries that served to reopen the question of Nestorius's orthodoxy. The first of these texts was Friedrich Loofs's edition of the *Nestoriana* (1905) in which he collected all of the previously known works of Nestorius as well as approximately 100 newly discovered fragments. The second important factor was the 1895 discovery of Nestorius's book called *Bazaar of Heracleides*, in a Syriac translation. This text was translated into English in 1925. In this text, Nestorius claimed that his words were misinterpreted, and that he believed that Christ had two natures, both human and divine, but in only one person. Scholars, however, are still divided as to whether Nestorius engaged in heretical rhetoric.

Historians began to examine the problem of Nestorius and his orthodoxy. The issues concerning Christ's divinity–humanity seemed to expand far beyond the *Theotokos* issue. Since there was serious contention between Alexandria and Constantinople throughout this time period, several scholars see friction as the real problem—both in the political realm and among the strong personalities of the era. For instance, Friedrich Loofs maintains that Cyril's interest in negating the teachings of Nestorius was a result of political maneuvering on Cyril's part, in order to avert attention away from his own teachings. G. L. Prestige argues that political and personal factors were certainly part of the problem, but that misunderstanding about language was a bigger issue (Prestige 1940: 264). According to Robert Sellers, both sides had a difficult time understanding the terminology used by the other side, but in essence were saying the same thing (1940: 208–14, 233).

More recently, scholars have examined the controversies and put forward the idea that while the differences in politics were significant, theological differences were more important. These historians argue that the fundamental issue centered on the view that Christ could only be depicted fully with a dual nature, according to boundaries set forth in scripture. Since the West adopts a use of concrete terms to describe the humanity of Christ, these academics generally view Western Christology to be closer to Nestorius's ideas than to Cyril's view.

A third group of scholars asks whether or not Nestorius believed that God was present at all in the incarnation, and to mixed results. Bethune-Baker posited in 1903 that Nestorius had taught two persons in Christ. However, after reading the *Bazaar of Heracleides* he changed his mind and agreed with Grillmeier that the Antioch bishops correctly identified the full humanity of Christ; but they did not go far enough in emphasizing the idea that God had become fully human (Bethune-Baker 1903: 274–76; Grillmeier 1975: 465–77). In contrast to these scholars, Donald Fairbairn argues that, on the one hand, the Nestorian dispute was not about the idea that there were two persons in Christ, or whether Christ was a single person, but instead the controversy was really about the identity of one personal subject. According to Fairbairn, Nestorius saw the *prosopon* of Christ as a unique hybrid of the divine and humanity. On the other hand, Cyril viewed Christ as a composite *prosopon*, as God himself who had become human and come to earth in order to provide humanity with a fuller communion. For Cyril, it did not matter how poorly Nestorius explained the composite notion of union in Christ, if there was a composite unity, then Christ was not divine and it might as well be the case that there were two sons of God (Fairbairn 2003: 130–32).

Conclusion

From the moment of Nestorius's excommunication until now there have been expressions of uncertainty as to whether he truly taught and believed what was later defined and condemned as Nestorianism (Braaten 1963: 251). The concept that became known as Nestorianism maintained that Christ was divided into to two persons, one human and the other divine, which denied the incarnation. The specific nature of Nestorius's belief system is still disputed, and until recently historians tended to view Nestorius from Cyril's point of view. Some scholars suggest that a mistranslation of the word *prosopon* caused most of the difficulties. The discovery of the text of the *Bazaar of Heracleides* has caused other scholars to reexamine the issues raised by the Council of Chalcedon and has convinced many that Nestorius was an orthodox theologian who was merely striving to negotiate a compromise between Antioch and Alexandria.

The theology of Nestorius continues to be problematic for historians. Nestorius was condemned and exiled as a heretic during the fifth century because of

his beliefs. After the Council of Ephesus in 431, however, the Assyrian church of the east (eastern Iraq and Iran) refused to condemn Nestorius as a heretic. This rejection eventually led to a split known as the Nestorian schism, which separated the Assyrian church from the Byzantine church. Nestorianism was condemned and eventually stamped out in the Roman world, but because of missionary activity it spread into the Arabian Peninsula, India, and China. There are a few Nestorian communities remaining today, mostly in the Arabian Peninsula, but also in India and the United States.

References

Bethune-Baker, J. F. *Early History of Christian Doctrine*. London: Metheun, 1903.

Bethune-Baker, J. F. *Nestorius and His Teaching: A Fresh Examination of the Evidence*. New York: Cambridge University Press, 1908.

Braaten, Carl E. "Modern Interpretations of Nestorius." *Church History* 32 (1963): 251–67.

Caner, Daniel. *Wandering, Begging Monks: Spiritual Authority and the Promotion of Monasticism in Late Antiquity*. Berkeley: University of California Press, 2002.

Cross, F. L. *The Oxford Dictionary of the Christian Church*. New York: Oxford University Press, 2005.

Evans, G. R. *The First Christian Theologians: An introduction to Theology in the Early Church*. London: Blackwell, 2004.

Fairbairn, Donald. *Grace and Christology in the Early Church*. New York: Oxford University Press, 2003.

Ferguson, Everett. *Encyclopedia of Early Christianity*. New York: Garland, 1997.

Frend, W. H. C. *Encyclopedia of the Early Church*, Vols. I and II. New York: Oxford University Press, 1992.

Grillmeier, Alois. *Christ in Christian Tradition*. Louisville, KY: John Knox Press, 1975.

Hastings, Adrian. *The Oxford Companion to Christian Thought*. New York: Oxford University Press, 2000.

Kelly, J. N. D. *Early Christian Doctrines*. London: Adam and Charles Black, 1978.

Loofs, Friedrich. *Nestorius and His Place in the History of Christian Doctrine*. New York: Cambridge University Press, 1914.

McGrath, Aliester E. *Christian Theology: An Introduction*. London: Blackwell, 2007.

McGuckin, John A. *St. Cyril of Alexandria the Christological Controversy: Its History, Theology and Texts*. Leiden: Brill, 1994.

Nystrom, Bradley P., and David P. Nystrom. *The History of Christianity: An Introduction*. New York: McGraw-Hill, 2003.

Olson, Roger E. *The Story of Christian Theology: Twenty Centuries of Tradition and Reform*. Downers Grove, IL: InterVarsity Press, 1999.

Pelikan, Jaroslav. *The Christian Tradition: A History and the Development of Doctrine*. Chicago: University of Chicago Press, 1971.

Prestige, G. L. *Fathers and Heretics: Six Studies in Dogmatic Faith with Prologue and Epilogue*. New York: Macmillan, 1940.

Rousseau, Philip. *The Early Christian Centuries*. New York: Longman, 2002.

Russell, Norman. *Cyril of Alexandria*. New York: Routledge, 2000.

Sellers, Robert. *Two Ancient Christologies: A Study in the Christological Thought of the Schools of Alexandria and Antioch in the Early History of Christian Doctrine*. London: Society for Promoting Christian Knowledge, 1940.

Urban, Linwood. *A Short History of Christian Thought*. New York: Oxford University Press, 1986.

Wessel, Susan. *Cyril of Alexandria and the Nestorian Controversy: The Making of a Saint and of a Heretic*. New York: Oxford University Press, 2004.

9

The Celtic Church that arose after 400 CE as distinct from Roman Catholicism is a modern construct, rather than a historical reality.

PRO	Michael Greaney
CON	Joseph P. Byrne

PRO

During the reign of Charles I Stuart (1600–1649), there were two "parties" in the Church of England. The Puritans, or "Low Church," rejected anything that hinted of "Roman" control or influence. The "High Church" preferred to retain the outward forms of Catholic worship, as well as most of the doctrines apart from papal supremacy. Puritans asserted that adopting the outward forms of Catholic worship made one a "papist," whether or not the worshiper acknowledged papal supremacy. This presented the Stuart state, actually three different kingdoms with wildly divergent interests and strong mutual antagonisms, with a serious problem. Beginning with James I (1566–1625), the aim of the king was not so much to rule well, but to rule at all. Religion being the most volatile issue, the Stuarts pursued a policy of general toleration among the four main religious groups: Catholics, Puritans, the Scottish Kirk (Covenanters), and the Established Church.

Charles I's idea of governance was to achieve unity as one of the highest priorities and to enhance royal power at the same time. Because religious differences were believed to be at the bottom of much of the political discontent, imposing uniformity of practice became crucial. In this, Charles was strongly influenced by William Laud, archbishop of Canterbury (1573–1645). It was Laud's views that would go a long way in convincing Charles that the Celtic Church was a separate and threatening religious body. Laud had a great attachment to the "externals" of Catholic worship, but had to tread very carefully with respect to doctrine, because the Puritans regarded Laud as a virtual "papist" on account of the Catholic forms he preferred and tried to impose on the Church of England. This came into conflict with Laud's goal of a unified, national church to support a unified state. Laud believed he could only unify the Church of England by imposing externals and a uniform prayer book. In Laud's eyes and his intellectual and religious successors', form became everything, substance nothing.

Laud's sophistry would reach its ultimate expression in the Oxford movement of the early 19th century and its development of "Branch theory." In the

175

eyes of Laud and his descendants, external practices, not adherence to specific doctrines, determine whether an organization is part of a universal church. Since the Celtic Church differed from the universal church on the continent on matters of form, it must, therefore, be a distinct and separate visible church, although part of the invisible church. The Catholic position (shared by the Orthodox and many Protestant churches) is that adherence to essential doctrines—most notably papal supremacy—makes a rite Catholic, not outward forms or liturgical practices. The issue, then, is whether the differences in liturgical practice that characterized the Celtic Church were "substantial" or merely "accidental," that is, involving only outward forms and not touching essential doctrine. The evidence of history supports the conclusion that the Celtic Church from the earliest time was, and always considered itself to be, in union with Rome. It was not a separate entity from the rest of the church in England, and thus, the idea of a distinct Celtic Church is a modern construct rather than a historically supported reality.

Palladius and Patrick

Contrary to popular belief, there were Christians in Ireland (called Scotia in late imperial and early medieval times) before the advent of Patrick, the "apostle of the Irish." Prosper of Aquitaine (ca. 390–455, 460, or 465, depending on the source consulted), recorded in his *Epitoma Chronicon* that Pope Celestine I (422–432) made deacon Palladius (ca. 408–ca. 460) a bishop and sent him on a mission "to the Scots believing in Christ" in the eighth year (431) of the reign of the Emperor Theodosius II (401–450).

The Palladius the pope commissioned as his first official representative to Ireland may be the same Palladius who recommended to Pope Celestine that Germanus be sent in 428 to dispute with adherents of "Pelagianism" in Britain where the heresy was particularly strong. Pelagianism denied original sin and Christian grace. Contemporaries disagreed as to Pelagius's country of origin. Augustine, Orosius, Prosper of Aquitaine, and Marius Mercator asserted that the cognomen "Brito" or "Britannicus" indicated that Pelagius was from Britain. Jerome, however (at odds with Augustine on most nondoctrinal matters), called Pelagius "a stupid fellow, stuffed with the porridge of the Scots," and claimed that the Irish diet affected the heresiarch's memory and reasoning power (Praef. in Jerem., Lib. I, III).

Jerome stated that Pelagius came from Ireland, staying in Britain only long enough to spread his doctrines there before traveling to Rome where he briefly succeeded in persuading Pope Zozimus of his orthodoxy. Augustine and Jerome put personal differences aside and convinced Zozimus to condemn Pelagianism.

The appointment of a bishop to an area outside the classical world bounded by the limits of the Roman Empire demonstrates the importance that Rome put on refuting and countering the heresy and in securing the orthodoxy of all

believers in matters of essential doctrine. The civil unrest prevalent at this time throughout the empire had serious effects on the church as well. Sending a mission to "the Gentiles" (as those who lived beyond the boundaries of the empire were termed) should have had a very low priority. There was a desperate need to provide bishops and priests to meet the needs of existing believers within the empire, with nothing to spare for spreading Christianity into new areas. Clearly, however, Pelagianism was considered so dangerous that all possible sources of Pelagianism were to be identified and the heresy extirpated. Consequently, the pope decided that the best and most effective response was to send official missions to both Britain and Ireland, thereby making certain that Pelagianism had no hidden base from which it could reemerge and endanger the church.

Unfortunately, Palladius was not equal to the task. As effective as he evidently was within the world of Romanized Britain and Gaul, Palladius seems to have lacked the necessary background for dealing with a people outside that particular milieu. The culture and society of Ireland were significantly different from that of the classical world and constituted what was, effectively, an alien environment. Within a year, Palladius was recalled and Patrick replaced him.

According to his "Confession," Patrick was born somewhere in Roman Britain, the son of a Decurian, a civic official, but was captured at an early age and sold as a slave in Ireland. After years in captivity, Patrick escaped and made his way to Gaul, where he studied for the priesthood and was ordained. Patrick claimed he heard the Irish calling to him in his dreams, urging him to return to Ireland and convert them to Christianity. It is not clear how Patrick was selected to replace Palladius, but his superiors were evidently aware that Patrick was the ideal candidate to send on a mission to the Irish.

Consequently, Patrick was named a bishop and sent to Ireland. He made a number of important converts almost immediately, and eventually set up an administration based on the Roman model, headquartered in the ancient cultic center of Armagh. Most importantly, Patrick, whatever his perceptions of his own inadequacies or lack of learning, seems to have been outstandingly successful at completely eliminating Pelagianism among the Irish Christians that he found and in inculcating orthodox Christianity among his converts.

Patrick's success seems to have been recognized in Rome for, according to the Annals of Ulster, Rome sent three "auxiliary bishops" to Ireland to assist Patrick in 439, and in 441 Pope Leo the Great confirmed Patrick as head of the church in Ireland. Possibly realizing that outward forms and customs are not as important as adherence to sound doctrine, Patrick was less successful in grafting the traditional administrative structure and liturgical practices of the Western Church onto the unique culture in Ireland.

Concerned more with preventing the spread of a heresy, Patrick would likely have been somewhat lax in building a foundation of support for outward forms in contrast to the effort he put into making certain the Irish were orthodox. For a

missionary operating in an alien environment, the important thing was the unity of belief in essential doctrines, not uniformity of outward practices.

Specific Differences

In the decades following Patrick's missionary effort, the Church in Ireland adapted to the unique conditions in that country. The most significant difference was that the Church in Ireland centered on monasteries instead of cities. The Irish tribal structure seemed particularly suited for this arrangement, with control of the local monastery often vested in the same family for generations. The abbot, not the bishop, became the most important administrative individual in the Irish Church. A bishop was doctrinally necessary in order to maintain the "apostolic succession" and the tie to the rest of the church. Politically and administratively, however, the bishop was of minor importance.

The style of "tonsure" was also different in the Irish Church. Tonsure is a rite in which a baptized and confirmed Christian is received into the clerical order by shaving all or a portion of his head. It was not universally practiced in the early church. Jerome disapproved of the practice. On the continent, the style of tonsure was adapted from that of slaves, whose heads were shaved in order to facilitate identification as social and legal inferiors. The tonsuring of a new cleric presumably symbolized the submission of the cleric as a slave of God. The practice was to shave the crown of the head, leaving a ring of hair. This led later commentators to suppose (erroneously) that the Western style of tonsure was in imitation of Jesus's Crown of Thorns.

In Ireland, however, monks were almost immediately esteemed as scholars and learned men, supplanting the Druids. Most authorities thus believe that the unusual style of Irish tonsure, in which the entire front of the head was shaved, leaving the hair in back to grow freely, was derived from a presumed Druidic tonsure. If true, this helped shift the veneration accorded to the old order of scholars and holy men to those of the new religion. Whatever the source, the style of tonsure practiced or lack thereof anywhere in the church did not affect doctrinal orthodoxy or the church's essential unity.

There were other variations that grew up in Ireland, such as from where in the sanctuary of the church the Gospels and other selections from the Bible were read, as well as the then-innovative practice of "private auricular confession" (i.e., confessing one's sins in private to a priest, instead of proclaiming them to the congregation). These, too, were administrative in nature and did not involve doctrinal matters. The practice of confessing sins in private was considered so beneficial in encouraging penitents to make good confessions that it was eventually adopted throughout the church.

The most important difference between the church in Ireland and the rest of the church on the continent, however, was in the method for calculating the date

for Easter each year. Ireland used a method of calculation introduced by Patrick, formerly used in Rome, while Rome continued to use the same method, but changed to a different cycle of years. Although nondoctrinal, differences in the calculation of the date for Easter often caused people with an inadequate understanding of the issue to accuse those whose practice differed from their own as being heretics or dissenters. That these and similar accusations are without foundation is demonstrated by the fact that differences in the calculation of the date of Easter were never considered an impediment to the admittedly transient reunions of the Western and Eastern Churches in 526, 681, 787, 869, 1274, and 1439.

Columbanus

The strongest evidence for the unity of the Irish Church with that on the continent is given by Columbanus, considered by many authorities to be the greatest and most influential monk from Ireland. Columbanus was born in Leinster in or about 540. Becoming a monk at an early age, it was not until he was about 35, in or about 575, that he requested permission from his abbot, Comgall, to go to Gaul. With a dozen companions (personally selected by Comgall both for experience and for the symbolism of the number 12) Columbanus established monasteries following Irish practices in Annegray, Luxeuil, and Fontaine.

Royal favor shown to Columbanus seriously undermined the support previously enjoyed by the native Gaulish bishops, many of which were members of the nobility and who, despite the prevalence of simony (buying and selling church offices) and other sins, were Columbanus's nominal superiors. A Gaulish Church Council had, a few years before Columbanus's arrival, enacted decrees such as: no monastery or hermitage could be founded without the consent of the local "ordinary" (bishop); no abbot could rule more than one community; each abbot had to report yearly to the local ordinary; no abbot could absent himself from his monastery, make important decisions, or accept gifts of landed property without the permission of his ordinary; and that the monks' fasts and liturgical practices had to be approved by the local ordinary.

Columbanus, coming from the Irish culture in which the bishop was a minor, if necessary figure, violated every one of these ordinances. More concerned with unity of belief and doctrine, he likely gave no thought to the possibility that he was deeply offending some very powerful political opponents.

Nevertheless, the Gaulish bishops had to be cautious in how they handled the Irish interloper. Columbanus enjoyed a significant measure of royal support, while the bishops' credibility, if we can believe the conditions described by Gregory of Tours (ca. 539–595), was virtually nonexistent due to the decay of the local churches under their control. The Gaulish bishops therefore took the extremely dangerous and, for them, questionable step of accusing Columbanus of heresy or, at least, schism.

The native bishops based their accusations on the fact that, in the matter of tonsure and the calculation of the date of Easter, Irish practice differed from that of Rome. The facts that tonsure was hardly a doctrinal issue and that the method of calculating Easter used in Gaul also differed from that of Rome were ignored. The value of these issues in the eyes of the Gaulish bishops was that they were in an area in which the civil authorities had no power, and thus they could not interfere.

Faced with this difficult situation, Columbanus appealed to Rome. He wrote a series of letters to Pope Gregory, presenting his case and requesting that the pope issue a judgment in the matter. This was a bold step that put a stop to the Gaulish bishops' accusations. The letters are not only masterpieces of strategy and diplomacy, but they contain clear and unequivocal statements that Columbanus considered the pope the head of a universal church, that the pope was the obvious judge in a dispute of this nature, and, finally, that the Irish Church was in full communion with the church on the continent, headed by the pope, and not a separate entity: "All we Irish, inhabitants of the world's edge, are disciples of Saints Peter and Paul and of all the disciples who wrote the sacred canon by the Holy Ghost, and we accept nothing outside the evangelical and apostolic teaching; none has been heretic . . . none a schismatic; but the Catholic Faith, as it was delivered by you first, who are the successors of the holy apostles, is maintained unbroken" (Columbanus 2008: *Epistola* III).

Columbanus's concern was more for the unity of the church than in preserving the unique Irish liturgical heritage. He expresses a willingness to abide by the pope's decision, but it is plain that the real issue is not a difference in liturgical practice, but the underhanded tactics employed by the Gaulish bishops to rid themselves of a political rival.

Columbanus's concern for the unity of the church expressed itself a few years later in another letter he wrote to Pope Boniface IV, after Columbanus's expulsion from Gaul in 610 due to his refusal to retract his condemnation of the grandson of Queen Brunhild, Thiery II of Burgundy, for licentious behavior and loose living. The letter resulted from the Irish monk's concern for the schism over the "Three Chapters Controversy." This was a complex and extremely esoteric argument rooted in disagreement over the interpretation of the writings of three eastern theologians. The argument over whether writers were or were not orthodox was tearing the church apart. Several areas went into schism for a number of years, while others were close to a break.

Columbanus admitted that he did not understand the controversy, but he considered the unity of the church paramount. A theological argument on such abstruse matters could hardly, in his opinion, be so important that it was worth destroying the church. He urged the pope to take the lead in settling the question and to bring the church back together. The pope, the head of the universal church, was the obvious person to act as judge and to settle the matter.

It is thus reasonable to conclude, based on the statements and beliefs of one whom many authorities consider the quintessential Irish monk and the strongest defender of specifically Irish liturgical practices, that the Irish Church was an integral, even critically important, part of the universal church, as the widespread missionary efforts of the following two centuries were to demonstrate.

Irish Missionary Effort

Many authorities credit the Irish missionary effort in Europe from the 6th to the 11th centuries not only with reviving Christianity in areas where it had decayed, but with preserving and spreading what remained of classical learning after the implosion of the classical Roman Empire and the shift of the imperial "regnum" (rule) to Constantinople.

Despite—or, possibly, because of—the liturgical differences that characterized the Celtic rite, Irish influence was pervasive throughout Europe. Comparing the rigor and asceticism of the Irish missionaries with the (possibly exaggerated) laxity of the local clergy (Gregory of Tours lists clerical crimes and failings in his *History of the Franks*), kings and nobles preferred the advice and counsel of Irish monks. They made grants of land and wealth, providing the financial means for the establishment of the vast number of monasteries for both men and women that flourished throughout the Middle Ages, many of which are still in existence today.

Taking account only of the major institutions, these Irish "foundations" on the continent ranged from Ghent and Köln in the north, to Vienna and Salzburg in the east, and Tarentum and Naples in the south, and included such notable centers of religion and learning as Fulda, Paris, Lexeuil, Saint Gall, Berne, Milan, Bobbio, and even Rome itself. All the *Schottenklöster* ("Scottish" monasteries) eventually adopted the Benedictine "rule" to replace that of Columbanus.

The shift from the rule established by Columbanus to that of Benedict was gradual, resulting in the "Iro-Frankish" tradition, and eventually complete integration into the regular practices of the church on the continent. There are no recorded instances of violence or rebellion resulting from the change, although local rulers were known to evict Irish-born monks, as Brunhild did to Columbanus and his companions, but leaving the native-born Irish-trained clerics in place.

Such expulsions, however, were clearly political acts by the civil authorities, not religious matters, doctrinal or administrative. The peaceful and gradual integration of the Irish foundations, together with the vast number of them, even in the center of power of the Western Church, Rome, offers convincing proof that the Celtic Church was never considered a separate establishment, however much it might have differed on nondoctrinal liturgical practices. The form of religious belief and practice was different, but the substantial nature of the universal church remained fully intact.
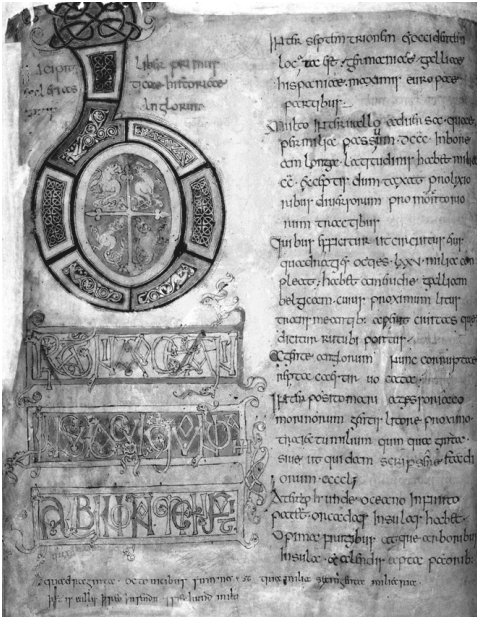
The Venerable Bede and the Synod of Whitby

In his *Ecclesiastical History of the English People* (1994), the Venerable Bede supplies strong evidence that the Celtic Church was an integral part of the universal church. In the approximately 54 references to the Irish Church and its unique liturgical practices, there are no hints that Bede considered the Irish, whom he clearly admired, anything other than orthodox and in full communion (union) with the universal church headed by the pope. Bede's concern was not heresy or schism, but that differences in practice might eventually lead to conflict between members of the same church.

Bede's account of the Synod of Whitby, in which the issue was settled for England, is revealing. As Bede relates, the kingdom of Northumbria at the time had corulers, each trained in a different tradition. Oswiu, the father, had been taught the Irish tradition out of Lindisfarne, while his son, Alhfrith, had been rigorously instructed in the Latin tradition by a tutor, Wilfrid of Ripon, trained in Rome itself. The court and the people were split in their observances, with the most obvious being the celebration of Easter. Half the people would be celebrating the risen Lord, while the rest were still keeping the Lenten fast.

A page from the Anglo-Saxon theologian Venerable Bede's *Historia Ecclesiastica Gentis Anglorum*, (A History of the English Church and People), completed around 731. His most famous work, the History traces the events from the time of Roman Britain through the establishment of Roman Christianity in England. (HIP/Art Resource, NY)

Matters came to a head when Ronan, an Irishman trained in the Latin tradition (whom Bede describes as a "violent defender of the true Easter," and "a man of fierce temper") caused the co-kings, Oswiu and Alhfrith, to request a synod in 664 to discuss the issues and decide on one, uniform practice for the kingdom. Because Northumbria exerted influence far beyond its own borders, the decision would determine which rite would predominate and, eventually, exclude the other throughout England.

King Oswiu opened the conference by stating its purpose: that all who served the one, true God should have a uniformity of observance, and that they were called together to determine which of the usages was the "truer tradition." Colman, selected to present the Irish argument, opened the

The Synod of Whitby: The Roman vs. the Celtic Church

The Venerable Bede, the great English church historian, recorded the proceedings of the Synod of Whitby, held in 664 CE, during which time the date of Easter was debated between those holding to the Roman Catholic tradition and those holding to the Celtic tradition. The debates may have centered on Easter, but had ramifications for the practice of Celtic Christianity.

"You certainly sin if, having heard the decree of the apostolic see, and of the universal Church, and that the same is confirmed by Holy Writ, you refuse to follow them; for, though your fathers were holy, do you think that their small number, in a corner of the remotest island, is to be preferred before the universal Church of Christ throughout the world? And though that Columba of yours (and, I may say, ours also, if he was Christ's servant) was a holy man and powerful in miracles, yet should he be preferred before the most blessed prince of the apostles, to whom our Lord said, 'Thou art Peter, and upon this rock I will build my church; and the gates of hell shall not prevail against it. And I will give up to thee the keys of the kingdom of heaven'?"

When Wilfrid had spoken thus, the king said, "Is it true, Colman, that these words were spoken to Peter by our Lord?" He answered, "It is true, O king!" Then said he, "Can you show any such power given to your Columba?" Colman answered, "None." Then added the king, "Do both of you agree that these words were principally directed to Peter, and that the keys of heaven were given to him by our Lord?" They both answered, "We do." Then the king concluded "And I also say unto you, that he is the doorkeeper, whom I will not contradict, but will, as far as I know and am able in all things obey his decrees, lest when I come to the gate of the kingdom of heaven there should be none to open them he being my adversary who is proved to have the keys." The king having said this, all present, both great and small gave their assent and, renouncing the more imperfect institution, resolved to conform to that which they found to be better.

Source: Bede. *The Ecclesiastical History of the English Nation*. London: J. M. Dent; New York: Dutton, 1910.

debate. Wilfrid followed, defending the observances of the Latin rite. Wilfred won the debate when he related Christ's institution of the papacy: "Thou art Peter and upon this rock I will build my Church and the gates of hell shall not prevail against it, and I will give unto thee the keys of the kingdom of heaven" (Matt. 16:18–19). Questioned by Oswiu, Colman admitted in effect that the popes, the heirs of Peter, were, in his opinion, the supreme authority in the church, whereupon Oswiu decided in favor of the Roman observances.

The terms of the synod and the manner in which the matter was settled clearly establish the fact that both sides believed themselves to be members of a universal church. The issue was whether unity of form should match unity of belief and doctrine, or whether such differences in form could be tolerated in the name of a

deeper unity. It was never a conflict between two separate churches, a concept that Bede, as well as the participants in the synod, would have found incomprehensible.

Charlemagne

When Charlemagne assumed the *regnum* of the western portion of the Roman Empire, he made immediate opponents of the Byzantine rulers who also claimed the right to rule the entire ancient territory once governed by Rome. To bolster his claim, Charlemagne needed the support of the pope who, in return for the protection given to him by the Frankish ruler, vested Charlemagne with the imperial crown on Christmas Day in 800. The need to collect allies against Byzantine claims is amply demonstrated by the embassy sent to the Frankish court by the legendary Haroun al Raschid (of "Arabian Nights" fame), which included among the gifts the first elephant seen in Europe in centuries.

While acclaimed as the ruler of a reformed Roman Empire in the west, Charlemagne was still very much a barbarian, albeit an extremely self-conscious one. Despite continuing efforts, he never learned to read or write, and he tended to rely on forcible conversion of recalcitrants and pagans to Christianity as a means of unifying his new empire. Charlemagne's "horrific" conversion of the Saxons on threat of extermination is an example of his slightly misdirected enthusiasm for political and religious unity.

Despite his personal illiteracy (or possibly because of it), Charlemagne had great respect for scholarship and sponsored what became known as the "Carolingian renaissance." To revive learning, Charlemagne imported monks from Ireland. Given Charlemagne's need to retain the support of the pope against the rulers of Constantinople and the Eastern Church that had their support, it is extremely unlikely that the straightforward and somewhat literal-minded new Roman emperor would have relied on monks from a rival church. This, while not conclusive, adds circumstantial evidence that the Celtic Church was, despite differences in liturgical practices, an integral part of the Western Church.

Malachy of Armagh and Bernard of Clairvaux

Two reasons are generally given to justify the Norman invasion of Ireland in or about 1169. One was to stop the slave trade between the western coast of England and the eastern part of Ireland. The other was to halt the decay of the Irish Church and reform it in order to bring it more into line with the continental norm.

Contradicting the alleged religious motives for the conquest, the holiness of the Irish clergy and the effectiveness of their pastoral care were known throughout the entire Western world. Bernard of Clairvaux, one of the strictest and most rigorous reformers of the medieval church, was greatly impressed with the Irish priesthood. His best friend was Malachy O'More, archbishop of Armagh,

Primate of Ireland, who died in Saint Bernard's arms and was buried in Bernard's habit. When Saint Bernard's time came to die, he was in turn buried in the archaic habit of the Irish archbishop.

These endorsements negate the claim that the Irish Church was in need of reform, or that the Celtic Church was a separate establishment from that of the rest of the church in Europe.

Laudabilitur

While the authenticity of *Laudabilitur*, the Papal Bull allegedly issued to Henry II Plantagenet by Pope Hadrian IV permitting an invasion of Ireland and a transfer of the temporal rule to the English Crown, has been called into question, the fact that the argument was used at all indisputably establishes the fact that the people of the 12th century—both Irish and English—believed the Irish Church to be under the authority of Rome.

The story is that Henry II, seeking to add the country to his domains, went to the pope with a proposal that he, Henry Plantagenet, be given a papal mandate to bring about civil and ecclesiastical reform in Ireland. According to the protestations of the English king, the condition of the island was such that drastic action had to be taken or absolute chaos would soon take over, to the detriment of civil order and the people's immortal souls. The pope then issued a Bull that granted Henry II the temporal lordship of Ireland, to be held in fief from the pope. In return, Henry was to effect the necessary reforms and also pay a "Peter's pence" tribute annually to the Holy See, one penny for every house in Ireland. Letters in the royal archives, purported to be from Alexander III, Hadrian IV's successor, make mention of the Bull and confirm its provisions.

None of this makes any sense unless it is accepted as a given that people of the time believed that the pope had such power, and that the Church in Ireland and the Church in England were both integral parts of the universal church.

The Norman Invasion

The Norman invasion of Ireland offers further proof that the Celtic Church was considered in union with the universal church. While the Normans carried out a reform of sorts, as presumably permitted by *Laudabilitur*, it was not based on any desire to establish uniformity of religious practices or raise the moral tone of the Irish clergy. Instead, it was a campaign to destroy native institutions, seen as strange and alien, and replace them with a more familiar liturgical tradition.

Geraldus Cambrensis (Gerald of Wales), a chronicler of the conquest, gave high praise to the Irish clergy in *The History and Topography of Ireland* (1983). He described at some length their virtues (especially chastity), their strict observance of all rules and regulations, as well as the rigor of their fasts and other austerities.

How the actual reform was carried out may be demonstrated by taking the activities of John de Courcy, Lord of Ulster, and his Manx wife Affreca as typical. A number of history books credit the pair with making several religious foundations, but fail to mention how they carried these out.

The Irish Church had an overwhelmingly monastic character. Where other churches concentrated on enriching the episcopacy, the Irish expended their wealth on the monasteries. Consequently, those institutions that had escaped the inroads of the Danes had been accumulating the donations of the faithful for centuries. Irish monasteries were a favorite Viking target, as their store of wealth was well known. What the Danes overlooked, the Normans would soon gather in. When the de Courcys located a richly endowed native Irish monastery in or near their territory, they would expel the Irish monks or nuns, confiscate all the moveable wealth, and attach the lands to their own desmene.

The de Courcys would then reestablish the monastery, endow it with a token amount of land, and staff it with a few Norman monks. The remarkable thing about these maneuvers is that the native Irish did not resist, but accepted them as if de Courcy had a right to do as he did. The Irish clearly did not discern any substantive change in religion, only in administration.

The Tudor Reformation

Matters were different during and after the Reformation. Unlike the situation following the Norman Conquest, the religious changes of the Reformation were widely regarded as affecting the substantial (doctrinal) nature of the Church in Ireland. "Defend the faith" became the rallying cry of the resistance to the English, which succeeded to some degree in uniting the Irish, both native and Norman, against a common foe, who was perceived as attacking the true church. This had not been the case in previous invasions, which were often seen as personal quarrels of those directly involved, resulting in a lack of unified efforts to drive out the invaders.

Conclusion

There thus exists a great deal of evidence, both direct and circumstantial, that the church in Ireland was founded by missionaries sent from the pope and has always maintained union with the bishop of Rome as the recognized head of a universal church. There is, on the other hand, no evidence to suggest, directly or indirectly, that the church in Ireland was ever construed as an independent body, whether its presumed foundation is traced to apocryphal missionaries from Egypt who founded a "Coptic-Celtic Orthodox" Church, or to Joseph of Arimathea who allegedly established Christianity in Britain after traveling to Marseilles with the apostle Philip, Lazarus, and Mary Magdalene, leaving Mary Magdalene to stay in Gaul.

References and Further Reading

Attwater, Donald. *The Avenel Dictionary of Saints*. New York: Avenel Books, 1965.

Bede the Venerable. *The Ecclesiastical History of the English People*. New York: Oxford University Press, 1994.

Bernard of Clairvaux. *The Life and Death of Saint Malachy the Irishman*. Kalamazoo, MI: Cistercian Publications, 1978.

Bieler, Ludwig. *The Works of Saint Patrick*. Westminster, MD: Newman Press, 1963.

Bieler, Ludwig, ed. *Four Latin Lives of St. Patrick, Colgan's Vita Secunda, Quarta, Tertia, and Quinta*. Dublin, Ireland: Dublin Institute for Advanced Studies, 1971.

Columbanus. *Letters of Columbanus*. Cork, Ireland: University College, 2008.

Columbanus. *Sancti Opera Columbani*. Dublin, Ireland: The Dublin Institute for Advanced Studies, 1970.

D'Arcy, Mary Ryan. *The Saints of Ireland*. St. Paul, MN: Irish American Cultural Affairs Institute, 1974.

Einhard and Notker the Stammerer. *Two Lives of Charlemagne*. London: Penguin Books, 1969.

Geary, Patrick J. *Before France and Germany: The Creation and Transformation of the Merovingian World*. New York: Oxford University Press, 1988.

Geoffrey of Tours. *The History of the Franks*. London: Penguin Books, 1974.

Gerald of Wales. *The History and Topography of Ireland*. London: Penguin Books, 1983.

Lonigan, Paul R. *The Early Irish Church*. Woodside, NY: Celtic Heritage Press, and the Aodh Ruadh Ó Domhnaill Guild, 1988.

Martin, F. X., and Moody, T. W. *The Course of Irish History*. Boulder, CO: Roberts Reinhart Publishers, 1967.

Reilly, Robert T. *Irish Saints*. New York: Avenel Books, 1964.

Scholz, Bernhard Walter. *Carolingian Chronicles: Royal Frankish Annals and Nithard's Histories*. Ann Arbor: University of Michigan Press, 1970.

CON

Christianity was an offshoot of Judaism that developed in the cultural and political matrix of the first four centuries of the Roman Empire. As missionaries carried the religion throughout the empire, they found converts among people of various classes and ethnicities, virtually all of whom had experience of Roman

life, imperial government, and Latin (or Greek) letters. Celtic peoples were scattered across the empire from Galatia (in Asia Minor) in the east to Wales in the west, and many of them were drawn to the gospel. By the later fourth century, Christianity dominated the Western empire, but the region itself slowly slipped away from Roman imperial control as pagan Germanic peoples forcefully migrated through Gaul, into Spain, and eventually into a Britain that had been largely abandoned by the Roman military after 410. Before the 430s, Irish people who had never known Roman rule became Christian, though nothing is known of this process nor of the contours of Christianity in Ireland before the arrival of Saint Patrick (ca. 432).

Cut off in many ways from the Christian sources of church administration and culture in the Mediterranean, Christian Celts, especially in Britain and Ireland, developed distinctive forms and practices that may have differed enough from those of the Roman Catholic Church to warrant labeling theirs a "Celtic Church." At distinct points in the 12th, 16th, 18th, 19th, and 20th centuries, historians and historically minded clerics have emphasized these differences. Yet there is not a shred of evidence that Celts of the era considered their church to be distinct from Christ's church (Jesus spoke of only one), though they (and others) clearly recognized often-important differences between themselves and *Romani*. In fact, from the early 1960s, historians of medieval Ireland and other Celtic areas have ceased discussing a "Celtic" Church altogether, emphasizing the differences among the Celt-Iberians, Welsh, Irish, Gauls, and Britons, rather than similarities that clearly distinguished them from emerging Roman Christianity or Catholicism. Dominating the concept of "Celtic," however, were always the Irish, the survival of whose records and whose peculiar position as a non-Romanized people had always made them the historiographical core of any "Celtic Church." And so today the term "Irish Church" generally replaces "Celtic Church" in genuinely historical discussions.

Most contemporary medievalists downplay the distinctiveness of even the Irish Church, however, either avoiding the label or qualifying it as not indicative of a church apart from that of Rome. In some ways, then, the controversy is settled, and not in the favor of this side. But historians are not the only ones with a stake in the matter. At various points in history, pointedly since the early 1960s and certainly since the early 1980s, some Christians and those drawn to Christian spirituality have discovered and explicated a distinctively "Celtic" Christianity or spirituality that they interpret as more "Jesus-like" and authentically Christian than Catholicism, Orthodoxy, or various expressions of Protestantism. Their books on Celtic Christianity or spirituality often play with very modern concerns over personal spiritual journeys, feminism, environmentalism, and patriarchal authoritarianism. Though harnessing historical figures, artifacts, writings, and events, modern Celtic Christians are far less interested in accurately interpreting the past than in providing a Christian alternative to traditional Christianity.

So, even discounting pan-Celtic and Celtic spirituality issues, the Irish Church in the early Middle Ages was distinctive enough from the Roman church on the continent to warrant calling it a separate church. In order to prove this point, we first need to examine the roots and expressions of that distinctiveness, and then provide a reasonable definition of "church" that allows for its application.

That the Celts Were Different

The Jesus movement, or Christianity, emerged out of first-century Judaism in a matrix of Greco-Roman culture set in an empire dominated by Rome at the height of its power. Its earliest missionaries spread the gospel among Jews and Gentiles in the Roman Empire, with only dimly understood efforts in non-Roman lands such as Ethiopia and India. As the movement developed into a church from the later first through fourth centuries, it adapted itself to the world Rome had built. Its holy scripture, or Bible, was available in Greek (both Old and New Testaments), with unofficial Latin versions preceding the "official" production of the Latin Vulgate by Saint Jerome in the late fourth century. Its organization evolved from small congregations huddled for worship in private house churches to a well-developed, hierarchical structure that borrowed freely from the declining Roman state. Regional leadership was provided by five patriarchs (father-rulers) who settled in the great cities of Constantinople, Antioch, Alexandria, and Jerusalem in the East and Rome in the West, while local administration was handled by bishops (overseers) and archbishops (leading overseers) whose seats (sees) were in the second tier of imperial cities and whose dioceses extended across the *civitates* (administrative districts) of the empire. These leaders had benefited from classical educations and often years of experience as civil or imperial officials. Christian intellectuals, or theologians, who had also benefited from secular Greek and Roman educations, developed intricate interpretations of the Hebrew and Christian scriptures as they sought to apply God's Word imbedded in them to social, political, cultural, and personal matters. They taught, wrote, and sometimes became bishops themselves.

Great stone basilicas built in the monumental Roman style replaced simple house churches, and these became home to broad fields of fresco and mosaics that pictorially proclaimed the Christian messages. Specially appointed (ordained)—and usually trained—priests aided the bishops or themselves led worship services or liturgies (including the Mass) in these grand spaces, using vessels and implements beaten and molded of precious metals and studded with jewels for the service and glory of God. At smaller local churches in towns and scattered across the countryside, and at shrines to heroes of Christian history (usually martyrs), priests also said the Mass, which combined the reading of scripture and preaching with the re-presentation of the Last Supper of Jesus and his apostles (the Eucharist with its bread and wine) as recorded in the New Testament and commanded by Jesus.

Along with teaching, preaching, and directing the liturgies, the priesthood, or clergy, was also responsible for an emerging set of rituals that bound the believer to the church and aided life in this world and the next. At birth or the point of formal conversion, the new Christian was baptized by either being immersed in water or by having water poured or sprinkled over his or her head by a priest or bishop, who also "sealed" the person with blessed olive oil (chrism). The spiritually purified new member of the church might over time, however, succumb to human weakness and Satan's temptations and sin by breaking God's moral law. For these people there was an evolving practice known as the Sacrament of Penance, which combined personal spiritual sorrow and repentance for disobedience to God and his church with a public, physical manifestation of that sorrow and desire to be forgiven in the form of actions ranging from prayer to pilgrimage. Christian rituals, or sacraments, also evolved around marriage and death and burial, and so Christianity came to envelope the believer from cradle to grave. The church proved a powerful mediator between innately sinful humans and a divinity who was seen as paradoxically both just and merciful. Like the emperor in the physical world, the Christian God wielded arbitrary power of spiritual (and eternal) life or death, focused in the act of final judgment and relegation of every person to the delights of heaven or the torments of hell.

Finally, both Eastern and Western Christianity developed monasticism for those who sought a spiritually focused life away from other social obligations. Beginning with hermits and later monastic communities in Egypt, this important institution spread north through Greek-speaking territories and north and west into North Africa, Italy, and Gaul. Eventually the sixth-century Italian hermit St. Benedict devised the rule (*Regula*) according to which most Western monasteries (Benedictine) organized and ran themselves.

It was Saint Paul himself who first preached and wrote to the Celts of Galatia in northern Asia Minor within a couple of decades of the Crucifixion. He had longed to go to Spain, whose inhabitants included many Celt-Iberians, though the gospel would arrive there in the hands of other missionaries. Celts of Gaul (Gauls) and of Britain (Britons) welcomed the new religion as its messengers sailed up the Rhone River and its sister streams and across the English Channel with merchants or soldiers. As part of the Roman Empire, all of these regions enjoyed regular commerce and communication with the Roman Mediterranean heartland. Latin, if often in a debased form, replaced or supplemented local Germanic and Celtic tongues, and some level of cultural integration aided Christianity's spread wherever Roman roads stretched. Without a doubt, Christian communities of Celtic Romans began to appear in the Western empire by the end of the first century CE.

The date of Christianity's first appearance in Ireland is lost to history, but surely there were Irish Christians before Bishop Palladius was dispatched to serve them in 431. The Romans had never controlled the island, nor had they ever tried. Pagan Irish people interacted with Christians in Britain, Gaul, and

probably Spain through trade, piracy, and perhaps in service as mercenaries. Saint Patrick was a Christian Briton who was captured by Irish pirates and served as a shepherd. Escaping, he returned to Britain, then traveled to Gaul, where he prepared for a clerical life. Although the date 432 is usually given for his return to Ireland as a missionary, this is uncertain. Unlike Palladius, Patrick spread the gospel among nonbelievers, establishing a lasting Christian presence in north-central Ireland.

Despite Patrick's successes, and whatever those of Palladius, Christianity remained a minority religion for quite some time, and pagan worship and culture remained visible for nearly two centuries. Pagan Celtic culture centered on the Druids, who served their society as priests, healers, and scholars. Their maintenance of ritual and sacrifices kept the many Irish gods content and promised success to the leaders of Irish society. The *filid* was both a spiritual seer and poet who traveled among Irish settlements, bringing news of both the physical and spirit worlds and entertaining high- and low-born alike. Usually called "*brehon* law," after the law-speakers who maintained the society's legal framework, Irish secular law and its processes were well developed. After the arrival of Christianity and Roman writing technology it was recorded along with church laws (canon law). The Irish people were organized into small kingdoms called *tuatha*, each of which was led by a king or *rí*, and kinship groups formed the basic building blocks of the *tuatha*. A warrior elite maintained its social position by raiding and fighting battles for the *rí*. It is impossible to speak of Irish towns before the 10th century, when Viking trading posts like Limerick, Cork, and Dublin began to evolve. The Irish lived close to nature in small-scale settlements that knew little of stone construction and nothing of monumental architecture. They did, however, possess skills and traditions in the decorative arts of metalwork and sculpture and recorded their thoughts in a rune-like written alphabet known as *ogham*, which consisted of one or more perpendicular and diagonal strokes arranged in sequence along a horizontal line, not unlike the teeth in a comb.

When Christianity established roots in this society it began a process of social and material change that eventually transformed Irish society, but in the process Christianity had to adapt in many important ways. What follows is a detailed list of some of the principal issues that set the Irish Church apart from that being hammered out in Rome.

How the Celts Were Different

Having never bent the knee to imperial power, the Irish seem to have experienced little contact of any sort with the earthly head of the Roman Catholic Church. Bishop Palladius may have been deacon to Pope Celestine I, but his mission seems to have originated in Gaul, not Rome. Patrick's mission had no roots at all in Rome and established no apparent ties to the apostolic see. In fact, the later

The Deer's Cry, or The Breastplate of Saint Patrick

Though its attribution to Saint Patrick is dubious, this most popular of Celtic hymns, dating from the late seventh or early eighth century, certainly reflects many of the themes of the Celtic Christian tradition.

I arise to-day:
vast might, invocation of the Trinity,—
belief in a Threeness
confessing of Oneness
meeting in the Creator(?).

I arise to-day:
the might of Christ's birth and His baptism
the might of His Crucifixion and Burial
the might of His Resurrection and Ascension
the might [of] His Descent to the judgement of Doom.

I arise to-day:
might of gradeso of Cherubim
in obedience of Angels
[in ministration of Archangles*]
in hope of resurrection for the sake of reward
in prayers of Patriarchs
in prophecies of Prophets
in preachings of Apostles,
in faiths of Confessors
in innocenceo* of holy Virgins
in deeds of righteous men.

I arise to-day:
might of Heaven
brightness of Sun
whiteness of Snow
splendour of Fire
speed of Light
sweiftness* of Wind
depth of Sea
stability of Earth
firmness of Rock.

I arise to-day:
Might of God for my piloting
Wisdom of God for my guidance
Eye of God for my foresight
Ear of God for my hearing

Word of God for my utterance
Hand of God for my guardianship
Path of God for my precedence
Shield of God for my protection
Host of God for my salvation
against nares of demons
against allurements of vices
against solicitations of nature
against every person that wishes me ill
far and near
alone and in a crowd.
I invoke therefore all these forces to intervene between me and every fierce
merciless force that may come upon my body and my soul:

against incantations of false prophets
against black laws of paganism
against false laws of heresy
against deceit of idolatry
against spells of women and smiths and druids
against all knowledge that is forbidden the human soul.

Christ for my guardianship to-day
against poison, against burning,
against drowning, against wounding,
that there may come to me a multitude of rewards;
Christ with me, Christ before me,
Christ behind me, Christ in me,
Christ under me, Christ over me,
Christ to right of me, Christ to left of me,
Christ in lying down, Christ in sitting, Christ in rising up
Christ in the heart of every person, who may think of me!
Christ in the mouth of everyone one, who may speak to me!
Christ in every eye, which may look on me!
Christ in every ear, which may hear me!

I arise to-day:
vast might, invocation of the Trinity
belief in a Threeness
confession of Oneness
meeting in the Creator.

Source: *The Irish Liber Hymnorum*. Ed. and trans. J. H. Bernard and R. Atkinson. London: 1898.

Book of the Angel (seventh or eighth century) claimed that an angel, rather than the pope or other bishops, was responsible for "bishop" Patrick's ordination. In what were or had been Roman territories, universal authority vested in an emperor or pope was understood and acceptable, but in Ireland this was an alien concept. As the great monastic missionary Columbanus related to one pope, in Ireland, Rome was only "great and famous" for "that chair" of Saint Peter the apostle. The Irish Church recognized the popes who served as Peter's successors as advisers and judges of last resort, as clearly outlined in several collections of Irish ecclesiastical law (canons). The *Liber Angueli* (book of the angel), which sought to support the authority of the bishop of Armagh over all of Ireland, makes the same statement, referring to the pope as merely "having authority over the city of Rome." According to an imperial rescript, the popes had authority over all Western bishops; though, of course, the imperial arm had never touched Ireland. In the mid-seventh century, Irish clerics sought guidance from Rome as to the correct date on which to celebrate Easter each year (see below), but over the next four centuries there is no evidence of any other appeal (true also of the "Celtic" churches in Scotland and Wales). Popes sent no representatives (legates) to Ireland, and no Irish bishop traveled to Rome for his pallium (a wool stole that symbolized his office). Only pilgrims traveled from Ireland to Rome to pray at the shrines of Saints Peter and Paul; in fact, *róm* in Old Irish came to mean "burial place." Saints were more powerful and respected than popes, as Columbanus made clear to Pope Gregory I around 600. The pope had tried to force Irish missions in Gaul to adopt the Roman calculation of Easter: "don't make us have to choose," Columbanus warned, between Gregory and Saint Jerome (on whose purported authority Irish custom rested), for to abandon Jerome would be heretical. It seems clear that while the Irish Church never repudiated the Roman pontiff in the way Protestant churches of the Reformation era did, it certainly drew on a very different model of hierarchical administration.

The Roman Christian bishop, with his urban seat, cathedral church, and platoon of clerics and bureaucrats, was modeled on the Roman provincial governor, a model unfamiliar to the Irish. He also traced his lineage back to one of Jesus's apostles through the ritual of ordination—a problem for "Bishop" Patrick. The earliest Irish Christians lacked such leaders (hence Palladius's mission), but after the 430s bishops appear, with the *tuatha* as their dioceses. This meant that the organizational church had a structure that was directly blended into that of Irish society, rather than running parallel to it. This was reflected in the facts that the clergy as a whole came to be treated in Irish law as a separate kin group, and that bishops had the same "honor price"—the penalty due for a transgression against the person—as the *rí* (and the *filid* and *brehon*). This resulted in what historian Dáibhí Ó Cróinín (1995) labels a "tribal church" as opposed to the diocesan Roman model. Although continental rules prescribed that three bishops ordain a new one, the evidence suggests that it was rare in the early Irish Church

for more than one to participate, a complaint of the English archbishops of Canterbury Lanfranc and Anselm as late as the early 1100s.

By the early 600s, Ireland saw the rise of the native monastic system (see below) and the concomitant decline in episcopal (bishop's) authority. The abbots (*coarbs*) who ran the monasteries came to manage Irish religious society as well, and bishops were relegated to sacramental functions such as ordinations of priests and other bishops. The evidence suggests that many bishops, if not most, came to be directly connected to the monasteries, and that some were even hermits, such as Cuthbert of Farne, or wanderers, an abuse complained about by Archbishop of Canterbury Theodore of Tarsus in the later seventh century. By the eighth century, Irish abbots, who were usually appointed members of the kin groups that patronized the monasteries, absorbed the managerial, administrative, governing, and disciplinary powers associated with continental bishops. Abbots who were not priests and had none in their monasteries needed bishops for saying Mass, hearing confessions, and baptizing infants, but many abbots were priests, and some had been raised from the episcopacy to the abbacy, a process opposite to that found among the Romans.

Irish abbots ruled monasteries that were peculiar in the Christian world, and uniformly so. The origins of Irish monasticism are unclear, but the practice as it emerges into the historical record by the sixth century is more closely related to that practiced in Egypt under the fourth-century rule of Pachomius than that found in much of Gaul or Italy. It may have derived directly from the travels of Athanasius, the fourth-century biographer of Saint Anthony, the archetypal Egyptian hermit. Interestingly, Anthony appears in Irish art, such as sculpted crosses, long before he does in Roman art. Monasteries were founded by the leaders of kin groups, and they and land donors retained the right to determine the abbots for generations, with the position being essentially hereditary. In a very real sense they were familial institutions, generally associated with lay settlements, and bishops had no jurisdiction over them. Local churches were administered from the monasteries and these could be strung out and geographically intermixed, forming "families" of the individual monasteries known as *paruchiae*. Scholars have found this system of quasi-familial organization to be related to the increasingly visible system of clientage, whereby weaker individuals sought more powerful ones to serve in return for protection. In society, this began to weaken ties of kinship, while in the *paruchia* system the more powerful kin groups gained in strength by their extended ecclesiastical associations.

In a society without towns, monasteries even replaced iron-age hill forts as geographic centers of political power. Following the plague years of the mid-660s, the bishop or abbots of Armagh, a seat that claimed Saint Patrick as its founder, contended with that of Kildare, associated with the estimable female Saint Brigid, for supremacy over the Irish Church in the sense that Canterbury had over the English. In so doing, the leaders at Armagh gathered a wide and

numerous string of associated churches and monasteries, the "*Paruchia* of Saint Patrick," though they did not achieve their goal. By the 700s much of Irish Christian life was organized by and around the monasteries, which had also become the centers of Christian culture, including education, painting, metalwork, biblical study, and explication, in a world still peopled—if ever more lightly—by pagans and their Druids. But this intertwining of monasticism and the wider society may also be seen as a secularization of monasticism. Feuding clans meant feuding monasteries, and Clonmacnois fought two violent battles with Birr in 760 and Durrow in 764; after its battle, Durrow counted 200 of its own dead. From 697 to 780 the leaders of the Irish Church held no general meetings (synods), a sign of the failure of any central ecclesiastical authority, whether internal or foreign.

Irish spirituality was heavily invested in its monks, who became renowned for piety, asceticism, and learning, even in the Greek and Latin classics. After the flood of pagan Germanic peoples in Gaul and Britain from the early fourth century came Irish missionary monks, like Saint Columban, who founded the monastery at Iona, and Columbanus, who founded a *paruchia* of monasteries across Gaul and into northern Italy. Some modern commentators who have studied the written remains of Irish monks find a strain of naturalism that is largely absent from continental monastic writing of the period. Roman Catholic monks, influenced by the thought of Saint Augustine and Plato, are considered to have been opposed to nature, finding it alien to the spiritual life, and to have sought to deny or suppress even their human nature, because they considered it damaged by sin. Irish monks, on the other hand, are thought to have considered God-created nature inherently good and welcoming, perhaps since sin only affects the human spirit. It is this "dignity of nature," as expressed in poetry and prayer, that modern environmentalists and New Agers find so attractive. Yet Irish monks are also known for their self-denying asceticism, which sometimes seems downright masochistic. Early Christian Ireland produced no known martyrs, and this may have heightened the tendency to self-sacrifice among

Iona Abbey on the Scottish Isle of Mull is a Christian pilgrimage site dating from 600 CE. The island was the site of the monastery established by Saint Columba, who actively converted Picts to Christianity during the sixth and seventh centuries CE. (Corel)

those who wanted to witness to Christ with a full measure. This may have been behind the impulse to self-imposed peregrination or exile spent in wandering and missionary work, an impulse rarely if ever found among continental monks.

It was the tendency to piety and asceticism that fueled the eighth- and ninth-century monastic reform movement known as Céli Dé (or Culdee: servants/clients/serfs of God). Begun at Tallaght monastery by Mael-ruain in the 750s, it remained rooted at Tallaght and Finglas monasteries, but spread across much of Ireland. Like later continental reform movements, the Culdees sought to isolate monks from secular contamination, emphasizing prayer, personal labor, and a hermetic ethos rather than the communal cenobitic one. It died out by the later 800s, perhaps under the pressure of the Viking raids.

Peregrination and asceticism in general were forms of self-sacrifice made in the light of the individual's sin and the need for propitiation of a just God. Pilgrimage and prayer, especially that recited while in very uncomfortable positions (e.g., with arms outstretched for hours [*crucifigium*] or standing barefoot on gravel or thorn branches) were well-recognized forms of penance and reflected in Christendom's first penitentials (books of recommended penance). In the developing Roman church, penance remained public and after Saint Augustine the church emphasized the believer's total dependence on God's grace freely given for forgiveness and ultimately for salvation. Augustine and other early theologians condemned the Celtic (Irish? Welsh? Briton?) theologian Pelagius who taught that human repentance and efforts to follow God's law were rewarded with the requisite grace. Though formally condemned by the Roman church in 418, Pelagian influences are clearly evident in Irish theology and biblical commentary. This may have been the impetus for stressing the technicalities of penance and of shifting it from a public matter to a much more private one of confession to a "soul friend" (*anamchara*) and penance being carried out in isolation. If spiritual and physical exertions could satisfy a just God, then exert themselves they would. There were even special monasteries (for example at Tiree and Hinba) for monks undergoing periods of imposed penance. While the Catholic Church eventually picked up the Irish personal confession and penance, Pelagianism remained a heresy, providing critics of the Irish Church a clear target.

The peculiarities of the Irish Church became an active issue not in Ireland or Rome, but in Britain, where Irish missionaries had planted many churches and monasteries. During the early seventh century the southern Anglo-Saxon church had been organized by Catholic bishops centered in and stemming from Canterbury. The two Christian cultures clashed along their borderline at a synod held at Streanoeshalch (Whitby) under the gaze of King Oswiu of Northumbria in 664. One issue was the way in which the monks shaved their heads (as a sign of their religious calling) known as the tonsure. The Roman Benedictines left a

ring of hair, while the Irish shaved from the forehead back, leaving the sides long. Though seemingly a minor matter of taste, the Benedictines demanded uniformity, but the Irish claimed that the Benedictine style was that of the arch-heretic Simon Magus and, therefore, utterly unacceptable.

The various methods of the annual dating of Easter, dated from the fact that Passover, which necessarily preceded Easter, was determined by the Jewish lunar calendar, as opposed to the Julian calendar solar method used by Christians, were not in synch. Early in the fourth century the Christian church decided that it would not follow the Jewish calculation but would determine it according to its own rules. These computational rules were complex, however, and resulted in three different systems. As in many matters, the Irish Church followed an older, 84-year cycle that they claimed had its roots in the Gospel of Saint John the evangelist. The flaws in this system were addressed in 475 by a Roman named Victorius of Aquitaine, who produced the "Victorian" calculation system, which proved to be more regular. Later, in 525, the Roman monk Dionysius Exiguus developed an even more refined system. Although the English monks pressed for the acceptance of the Dionysian tables at Whitby, it is not clear that Rome itself had accepted them in place of the Victorian. It is clear that Oswiu accepted the "Roman" tonsure and dating system for his churches, while the Irish were left with theirs, a further symbol of their independence (though in fairness, the Irish had consulted with the pope over the appropriate system).

Modern proponents of a "Celtic Christianity" often stress that the Irish viewed Christ as far more personal and close to humankind than Roman Christians did. They note the rationalism of the Greco-Roman culture, the transcendence of the Roman Emperor (a convenient mental model for the divine), and the monumental churches with their apse mosaics of Christ as ruler and judge. For the Irish, Christ was *ard-rí*, High King, in a society where kingship was local and approachable. Graydon Snyder (2002) finds that the Celtic "I" replaces the Roman, communal "we" as Christians approach their God in prayer and meditation, and that the Jesus who rewards effort and piety directly challenges the far more distant and ineffable God of the Romans.

Irish church structures were often very small, and some believe that the Eucharist was actually prepared in the structure while the congregation stood outside; a far cry from the great basilicas and churches of the continent. Services seem to have been in Latin, but from an out-of-date version of scripture, the *Vetus latina* rather than Saint Jerome's *Vulgate*, and followed the older Gallican liturgy rather than the newer Roman forms. Other historians have stressed the degree to which Irish canon law is intertwined with secular and political concerns, a development far less noticeable among "Roman" collections. The distinctive nature of Irish or Hiberno-Saxon art, with so little classical influence, is also considered a distinguishing feature of Irish church culture.

Traditions of an Irish Church

Bede, the eighth-century English Benedictine who related the issues and out-
comes of the Synod of Whitby, may be said to be the earliest historian who rec-
ognized an Irish Church with characteristics distinctive from the Roman. These
were noted again by reform-minded clerics in Rome and England in the 11th
and 12th centuries, who sought to bring all Christian outliers under Rome's
direct control, a move that lost the orthodox churches to the Schism of 1054.
Irish ecclesiastical leaders complied, such that an Irish Church may be said to
have disappeared by ca. 1100. Even so, the English Pope Hadrian (Adrian) IV
recognized enough differences to allow a reforming "crusade" by the English
around 1170. Interest in the early Irish (Celtic) Church was revived with the
Anglican Reformation of the 16th century. Anglican Reformers who had split
from Rome saw the early Celtic Church as purer than its Roman counterpart,
and thus a model for the reformed one in England and Ireland. Archbishop of
Canterbury Matthew Parker wrote in 1572 of the superficiality and vain cere-
monies brought by his predecessor Augustine of Canterbury in 600 to the "pure
and immaculate" British (Celtic) church. For these clerics, the Celtic Church
was surprisingly Anglican. Eighteenth-century Romantics who sought the purer,
noncivilized roots of Western culture in the "folk" of the past and present
began interpreting the remains of the Irish Church in cultural (more primitive is
better) terms rather than confessional ones. In the later 19th and early 20th cen-
turies, it was Irish nationalist scholars like Douglas Hyde who found the distinc-
tively Irish spirit in literary and spiritual works of early medieval Ireland. For
them, the Irish Church was distinctive from that of the despised English, and
from that of the overly clerical Catholic Church whose hierarchy often clashed
with the nationalist aspirations. Finally, Celtic Christianity emerged as a roman-
tic alternative to Catholicism or Protestantism in the wake of the early stages of
religious feuding in Ireland. In the early 1960s, publication of a cheap version
of Alexander Carmichael's collection of Scots folklore, *Carmina Gadelica*
(Gaelic Songs) fueled popular interest in both pagan and Christian Celtic antiq-
uity, and a fascination with Celtic Christian spirituality has grown up alongside
neo-Druidism, neo-paganism, and Wicca. It was also the early 1960s when me-
dieval scholars began to downplay the differences between Celts and "Romans"
and emphasized the similarities between the two.

Conclusion

Clearly, differences between the Roman Church and the Church in Ireland were
substantial. Whether they were distinctive enough to define two churches is a
matter of how one defines "church." Since the Reformation Westerners have
become used to accepting multiple Christian churches and have ceased (for
the most part) from hurling changes of heresy and schism. Early Christians, on

the other hand, recognized that Christ had established a single church, and all other facsimiles—such as Gnostics or Arians—were something else. If one can speak today, however, of a Lutheran church, a Presbyterian church, and a Methodist church and see the differences (and similarities) as something other than historical, then recognizing the Irish Church as distinct from that which developed under the direction of Rome and the "Romans" is eminently reasonable.

References and Further Reading

Bitel, Lisa M. *Isle of the Saints: Monastic Settlement and Christian Community in Early Ireland*. Ithaca, NY: Cornell University Press, 1990.

Bradley, Ian. *The Celtic Way*. London: Darton Longman and Todd, 2003.

Cahill, Thomas. *How the Irish Saved Civilization*. New York: Doubleday, 1995.

Charles-Edwards, Thomas. *Early Christian Ireland*. New York: Cambridge University Press, 2000.

Corning, Caitlin. *The Celtic and Roman Traditions: Conflict and Consensus in the Early Medieval Church*. New York: Palgrave, 2006.

Cronin, Deborah. *Holy Ground: Celtic Christian Spirituality*. Nashville, TN: Upper Room Books, 1999.

Davies, Wendy. "The Myth of the Celtic Church." In *The Early Church in Wales* (pp. 12–21). Edited by Nancy Edwards and Alan Lane. Oxford: Oxbow, 1992.

Dáibhí Ó Cróinín. *A New History of Ireland*, Vol. 1. New York: Oxford University Press, 2005.

Dáibhí Ó Cróinín. *Early Medieval Ireland: 400–1200*. New York: Longman, 1995.

De Paor, Máire and Liam. *Early Christian Ireland*. New York: Thames and Hudson, 1958.

De Waal, Esther. *A World Made Whole: The Rediscovery of the Celtic Tradition*. London: Fount, 1991.

Friesen, Milton J. "Monasticism in Fifth to Seventh-century Ireland: A Study of the Establishment of Christianity in Irish-Celtic Culture." *Religious Studies and Theology* 23 (2004): 79–98.

Gougard, Louis. *Christianity in Celtic Lands*. Dublin: Four Courts, 1992 (orig. 1932).

Herren, Michael W., and Shirley Ann Brown. *Christ in Celtic Christianity: Britain and Ireland from the Fifth to the Tenth Century*. Rochester, NY: Boydell, 2002.

Hughes, Kathleen. *The Church in Early Christian Society*. Ithaca, NY: Cornell University Press, 1966.

Hughes, Kathleen. *Early Christian Ireland*. Ithaca, NY: Cornell University Press, 1972.

Hughes, Kathleen, and Ann Hamlin. *Celtic Monasticism. The Modern Traveler to the Early Irish Church*. New York: Seabury, 1981.

Hull, Eleanor. *Poem Book of the Gael*. London: Chatto and Windus, 1913.

Jestice, Phyllis G. *Encyclopedia of Irish Spirituality*. Denver: ABC-CLIO, 2000.

Joyce, Timothy J. *Celtic Christianity: A Sacred Tradition, a Vision of Hope*. Maryknoll, NY: Orbis, 1998.

Kerr, W. S. *The Independence of the Celtic Church in Ireland*. London: SPCK, 1931.

Low, Mary. *Celtic Christianity and Nature*. Edinburgh: University of Edinburgh Press, 1996.

Mackey, James. *An Introduction to Celtic Christianity*. Edinburgh: T&T Clark, 1989.

Mac Niocaill, Gearóid. *Ireland before the Vikings*. Dublin: Gill and Macmillan, 1972.

McNeill, John T. *The Celtic Churches: A History A.D. 200 to 1200*. Chicago: University of Chicago Press, 1974.

Meek, Donald E. *The Quest for Celtic Christianity*. Edinburgh: Handsel Press, 2000.

Moorhouse, Geoffrey. *Sun Dancing: Life in a Medieval Irish Monastery and How Celtic Spirituality Influenced the World*. New York: Harcourt Brace, 1997.

Mytum, Harold. *Origins of Early Christian Ireland*. New York: Routledge, 1992.

O'Loughlin, Thomas. *Celtic Theology*. New York: Continuum, 2000.

Olsen, Ted. *Christianity and the Celts*. Downers Grove, IL: InterVarsity Press, 2003.

Pierce, Susan. *The Early Church in Western Britain and Ireland*. Oxford: British Archeological Report, 1982.

Rees, R. B. *Pelagius: Life and Letters*. Rochester, NY: Boydell Press, 1998.

Roy, James Charles. *Islands of Storm*. Dublin: Dufour, 1991.

Scherman, Katherine. *The Flowering of Ireland: Saints, Scholars and Kings*. New York: Little, Brown, 1981.

Sellner, Edward. *The Wisdom of the Celtic Saints*. Notre Dame, IN: Ave Maria Press, 1993.

Snyder, Graydon. *Irish Jesus, Roman Jesus: The Formation of Early Irish Christianity*. Harrisburg, PA: Trinity Press, 2002.

10

The inhabitants of Easter Island who erected the monoliths were from South America, not from Polynesia.

PRO   Chris Howell
CON   Harald Haarmann

PRO

Thor Heyerdahl's 1950 assertion that Easter Island (Rapa Nui) was also settled by pre-Columbian South Americans was based on his archaeological work on both the island and mainland South America as well as on his experimental voyages on the *Kon-Tiki* rafting expeditions. While much of his data are still utilized by researchers today, most of his conclusions about prehistoric civilizations in Polynesia and South America have since been discarded based on more recent research that points to a primarily Polynesian settlement of Easter Island. However, new archaeological findings have researchers revisiting Heyerdahl's general idea that contact indeed took place between Polynesia and South America, with Easter Island a likely central point between the two regions.

Two recent archaeological finds have forced researchers to reevaluate when Easter Island was settled and what regions Easter Islanders were in contact with. The first involves new radiocarbon dating on initial human activity on Easter Island's main canoe landing area at Anakena beach to around 1200 CE, or almost 800 years later than the normative view held based on previous radiocarbon dates from bulk organic material in the three volcanic cones of the island. Archaeology professor Terry Hunt (2007) conducted the radiocarbon dating and excavations at Anakena beach while leading a University of Hawaii at Manoa research team in 2006.

The second find involves the excavation, radiocarbon dating, and DNA testing of Polynesian chicken remains from the pre-Columbian archaeological site of El Arenal on the Aracao Peninsula in Chile. These were discovered and tested by Alice Story and Elizabeth Matisoo-Smith, a biological anthropologist from the University of Auckland in New Zealand. The DNA tests reveal genetic ties to Polynesian chickens from Tonga, and the radiocarbon dating placed those remains in South America between 1300–1420 CE. These chickens may be related to the Araucana chicken of South America known on both Easter

203

The Columbian Exchange Reaches Easter Island

One of the most insightful historical arguments relating to the Americas during the 1970s and 1980s was the rise of what Alfred W. Crosby coined the "Columbian Exchange," or the biological relationship between Europe and the Americas. Crosby later expanded his theory worldwide, showing that the flora and fauna of the continents were transported across the oceans, often with unforeseen ramifications. In his 1986 book, Crosby explained his theory this way:

> Back in the Old World, most particularly in the densely populated areas of civilization, many organisms had taken advantage of contiguity with humans and their plants and animals to become their parasites and pathogens. These freeloaders often were slower to emigrate to the Neo-Europes than were humans and the organisms that humans intentionally brought with them. For example, Europeans brought wheat to North America and created the first of their several wheat belts in the Delaware River valley in the eighteenth century where the plant thrived in the absence of its enemies. Then its old nemesis, the Hessian fly, unjustly blamed on George III's mercenaries, who supposedly brought it across the Atlantic in their straw bedding, arrived and obliged farmers of the valley to find a new staple. (1986: 281)

This idea plays into the debate about the native people of Rapa Nui in that the existence of sweet potatoes and chickens on the island seems to defy the traditional Polynesian explanation, as both were not native to the Pacific islands, but rather to the Americas.

Source: Alfred W. Crosby. *Ecological Imperialism: The Biological Expansion of Europe, 900–1900.* Cambridge: Cambridge University Press, 1986.

Island and South America but not necessarily from Eurasia. A debate over the research data has emerged in 2008 in the journal of the *Proceedings of the National Academy of Sciences*, with Story et al. (2007) answering critics satisfactorily so far concerning issues of dating and genetic testing interpretations.

The presence of Polynesian chickens in a pre-Columbian archaeological context may help answer one of the great mysteries associated with eastern Polynesia. How did the world's greatest seafarers set out on purposeful expeditions and find some of the most remote islands in the world, especially Easter Island and the Hawaiian Islands, and yet miss the Western Hemisphere or the Americas entirely? Maybe they didn't. And if so, this calls into question not only the nature but also the directions of such contact, assuming the chickens did not engage in maritime navigation.

Taken together, these finds challenge the traditional questions of when Easter Island was settled and who the Easter Islanders were in contact with. Further, such finds call into question the nature and direction of previous research

and normative interpretations of Easter Island history. It may become necessary for the academic and research community associated with Easter Island to redirect questions, hypotheses, and research away from issues of earliest or first and toward the nature of human behaviors like contact and exchange if we are to become a more informed community on Easter Island human history.

Perhaps the basic questions, though not necessarily the conclusions developed by Thor Heyerdahl so long ago, are not dead at all. Was there contact between the Polynesian and South American worlds? If so, what was the nature and direction of that contact? How was Easter Island or Hawaii involved? Because so much evidence is missing from the past, can the research community really ever rule out the presence of South American Amerindians in Polynesia or vice versa? Are singular explanations and cause and effect really useful models for a better understanding of the past?

The idea of singular explanations for complex interactions between humans and environment is again at the forefront of research into the Easter Island mysteries. Researchers like Jerod Diamond (2005) have popularized an "ecocide" model for Polynesian settlers on Easter Island but have been contradicted by researchers such as Terry Hunt (2007), Paul Rainbird (2002), and Alice Story and her colleagues (2007) who suggest a more complex explanation is at work concerning human–environmental interactions. Essentially Diamond suggests humans were the problem in terms of collapse of Easter Island civilization while Hunt, Rainbird, and Story et al. suggest the Polynesian rat and European contact greatly complicate affairs for those interpreting human history associated with the island. According to the normative model, only Polynesians settled the island, only their impact on the environment created the ecocide, and only European historic descriptions later of a desolate island should be utilized to support that view. However, Hunt points out that early European seafaring visits to the island on Easter of 1722 CE by the Dutch captain Jacob Roggeveen actually describe it as lush and productive, with bananas, potatoes, sugarcane, chickens, coconuts, and remnants of decaying palm forests still visible. Archaeological and oral history data, however, suggest the island had undergone a human catastrophe by that time. Rainbird suggests that the arrival of European diseases, flora, and fauna, as was the case on other Polynesian islands, more likely created the ecocide. What are we to make of the contradictory interpretations and evidence? Perhaps that is just the nature of Easter Island research, a far more complex and contradictory world than singular interpretations might suggest.

Hunt and Orliac (2000) have both noted that linguistic evidence and radiocarbon dates from organic material in the craters were used to support an early occupation model of Easter Island between 300–800 CE. However, dated archaeological contexts definitely associated with human activity all point to a much later human presence and impact on the island, probably between 1200–1650 CE. It was the vague early dating evidence that lead Thor Heyerdahl to postulate that possibly two groups, with one from South America, helped settle Easter Island. Now that

new dating strongly suggests a later arrival, at least for Polynesians and the Pacific rat, we need to rethink much of the island's interpreted history. Such redating of Polynesian colonization of eastern Polynesia includes Easter Island, and the Hawaiian, Marquesas, Society, Austral, and New Zealand islands. In other words, researchers are now reinterpreting the historical context of Polynesian settlement all across eastern Polynesia and not just at Easter Island. The arrivals of Polynesians are later, usually around 1200 CE, and the interaction suggested between islands is far more planned and more frequent than had previously been thought. The same may be true for interactions with the Americas. Before examining these reinterpretations and their ramifications in reference to the general questions explored by Thor Heyerdahl concerning Rapa Nui or Easter Island, a brief review of some relevant data associated with Easter Island is in order.

Easter Island (Rapa Nui)

Easter Island is a remote volcanic landform of 171 square km in the Pacific Ocean over 2,300 miles from South America and over 1,200 miles from its nearest Pacific Island neighbor, the Pitcairn Islands. Before human arrival, it was rich with trees, birds, and marine species that could help support human settlement. Fresh water was available though not plentiful in the three volcanic craters of Terevaka, Poike, and Rano Kau, in the water table, and in low tide freshwater springs along the coast. Early Easter Islanders relied heavily upon local, wild animals for their food supply, such as porpoise and sea and land birds, in what Jerod Diamond describes as the richest bird species nesting ground in the eastern Pacific. Canoes were made from giant Easter Island palms (*Paschalococos disperta* or *Jubaea* subspecies), similar to the Chilean palm (*Jubaea chilensis*), which were utilized both as a food supply with palm hearts and as a raw material for tools. Two other giant trees *Alphitonia zizyphoides* and *Elaeocarpus rarotongensis* grew to 100 feet tall and to 50 feet tall, respectively. Both were ideal candidates for constructing the large canoes utilized by Polynesian seafaring expeditions. Dransfield suggests there may be interesting connections between the Easter Island palms and those in Chile but believes the paleopollen and seed evidence on Easter Island is not sufficient to explore the nature of those relationships as they might address human contact between the two areas.

Human Impact

At least 21 species of trees and 31 species of bird (including six land bird taxa) went extinct after human arrival by a combination of human settlement and introduced flora and fauna species, especially Pacific rats (*Rattus exulans*) and foodstuffs, according to Diamond. Important foodstuffs introduced by humans included the sweet potato (*Ipomoea batatas*), the chicken (*Gallus gallus*), and bananas, sugarcane, coconuts, taro root, and the ti bush, though droughts and salty winds are

Moai on Easter Island. The staggering architectural achievement of the people of Easter Island was the creation, especially the transportation and erection, of hundreds of *moai* monoliths. (iStockPhoto)

problematic for much of the Polynesian agricultural plants. Interestingly rat bones on Easter Island are far more common in trash midden sites than are fish bones. This suggests, according to researchers Barnes, Matisoo-Smith, and Hunt (2006), that rats may well be an intentional food resource for Polynesian expeditions.

Fascination with the Easter Island civilization that built the more than 800 monumental carved, stone heads (*moai*) has led to much scholarly debate and speculation as to who founded and settled it over 1,000 years ago. Traditional interpretations have favored the Polynesian seafarers from the western Pacific as the first and only settlers, while a few researchers associated with Thor Heyerdahl believe an additional group of South American origin also settled on Easter Island. Research in the past decade has focused more on the ecocide issue on the island. A few researchers such as Heyerdahl at Tucume in north Peru and Matisoo-Smith and Storey at El Arenal in Chile have been looking for South American connections with Easter Island.

Traditional Interpretations

The normative interpretation of Easter Island history based on academic research was that Polynesian seafarers who were spreading across the Pacific Ocean from

west to east over the past two millennia settled Easter Island. New Zealand, Easter Island, and the Hawaiian Islands were the last places reached as they were so far east and distant, toward the Americas or Western Hemisphere. Archaeological information, including radiocarbon dating of excavation layers, indicated that Easter Island was first settled by 400 CE. Easter Islanders settled the small island and lived in balance with the environment for perhaps 500 years or more, utilizing the island palms, dolphin, turtle, and marine resources, three freshwater volcanic lakes, and colonizing resources brought along in the outrigger canoes such as sweet potatoes, chickens, and maybe pigs and breadfruit. Then, almost suddenly, they began construction of monumental stone heads, not unlike those found on some other Polynesian occupied islands. This was followed by deforestation and soil erosion, loss of navigational ability, and eventually internal warfare that left the island civilization devastated and isolated until European ships arrived in 1722 CE and again in 1786 CE.

Critique of Traditional Interpretation

Interpreting human history even with well-researched and thought-out historical data is often a difficult proposition. Easter Island is certainly such a case. When the first Europeans came ashore in 1722 CE, the Dutch captain Jacob Roggeveen noted islanders of several classes and skin colors as did Captain Cook on a 1774 CE visit. Modern genetic testing of the few Rapa Nui left in modern times (Easter Island's indigenous inhabitants) revealed the presence of Polynesian ties. Genetic evidence was derived from select samples of modern Rapa Nui and also from burials in one of the prehistoric platforms on the island. So, clearly a strong Polynesian presence is associated with Easter Island both in the past and present. However, as human genetics researcher Cavilli-Sforza and his colleagues (1993) have noted, the limited sampling and strong evidence of historic bottlenecking or mixing of Easter Island populations due to slavery, disease, and deportation means that Thor Heyerdahl's 1950 hypothesis of Amerind genes influence cannot be rejected. Cavilli-Sforza et al. also note the unusual distance Easter Islander genetic data has with all other regions as compared even to other eastern Polynesians probably due to a founder effect and historic mixing. For instance, the Easter Islander genetic dataset's average distance from select South American and Southeast Asian groups is numerically 1,511 but other eastern Polynesians average genetic distance from those groups is only 1,031. As Cavilli-Sforza et al. suggest, Easter Island genetic history is greatly skewed due to historic processes and mixing. Without more sampling of the ancient Easter Islander remains we cannot be certain of past genetic relationships, especially if samples include modern islanders and migrations to the island.

Even the often-cited work by Erika Hagelberg (1994) that DNA extracted from 12 Easter Island skeletons from one platform shows the closest

relationship with Polynesians is open to critique. There are numerous platforms on the island, and sampling only one can hardly be representative of the multiple groups noted by Roggeveen in his 1722 CE visit to the island. The central Mexican site of Teotihuacan in Mesoamerica is a classic example of why this sampling can be problematical, though certainly understandable when dealing with the limited resources of archaeology. It was not until full-scale excavations by archaeologist Renee Millon (1966), that the multiethnic nature of the densest city in the pre-Columbian Americas was revealed. The planned city in central Mexico contained a number of trade barrios that indicated populations from and exchange networks with other Mesoamerican peoples, including distant Oaxaca and the Mayan city–state of Kaminaljuyu. Further explorations at Kaminaljuyu confirmed the strong ties, including marriage alliances with Teotihuacan. Had a sampling strategy been employed, many neighborhoods at Teotihuacan, comprised of local populations, would be the most likely sampled, and the connections with the rest of Mesoamerica might not have been fully understood. We must be careful in the case of Easter Island to avoid the assumption that because it is a small island, we somehow have a firm grasp on its ecological and human history. Current research certainly suggests otherwise.

Of course we cannot know or even sample a majority of the original Easter Islander inhabitants and thus could never conclusively test all past Easter Islanders for ties to their ancestral lands before arriving on Easter Island. Cavilli-Sforza et al. suggest this simple fact makes it difficult to ever conclusively answer the origins question from past genetic samples, though we can feel confident that a significant component is Polynesian in nature. But what about modern Rapa Nui? Why are modern Rapa Nui on Easter Island so problematical in genetic studies? Just looking at historic estimates of population gives us a clue. Captain Cook in 1774 CE estimated about 700 individuals on the island. The Spanish expedition of 1770 CE mentions between 900–3,000, La Perouse in 1776 estimated 2,000. Modern researchers also oscillate. Prehistoric estimates from Diamond could be as high as 30,000, from Hunt 6,000, and as low as 500 from Cavilli-Sforza et al. In 1862 slavers took 1,000 Easter Islanders to Peru where 900 died. Only 15 made it back to Easter Island, albeit bringing back with them smallpox or cattlepox. By 1877 only 110 Easter Islanders remained! What are we to make of such fluctuations in genetic and population data? That Easter Island history is indeed a complex affair not easily interpreted or understood, either by Heyerdahl or by modern researchers. Interestingly the 90 percent die-off rate of the islanders taken to Peru is very similar to that noted for Native American communities from the time of the Columbian exchange, though it should be noted that western Pacific peoples also suffered from introduced diseases at lesser rates.

The typical story of European colonial interaction on Easter Island led to the death of most Rapa Nui and mixed genetic background for the 110 who survived into modern times and genetic testing, as Rainbird has argued. This would call

into question the assumptions built into the early historic observations of class and caste as well as the validity of modern genetic studies. Just who were the Easter Islanders in the 1700s? How much interaction was already taking place with the outside world? What was the genetic background of the modern Easter Islander who were tested? Even the Easter Island script, known as RongaRonga, appears to be generated in the historic contact period as no evidence of the script can so far be dated in prehistoric archaeological contexts. In other words, historic impact on Easter Island has so skewed the record of the anthropological present and even the archaeological record via looting and disturbance, it is difficult to feel confident about research on Easter Islanders when it is applied to the past.

New Evidence

New evidence is emerging, however, that sheds further light on the prehistoric Eastern Pacific world and maybe on Easter Island itself. Both are archaeological finds. First, University of Hawaii at Manoa excavations into the island's only sand dune system and major canoe landing beach have produced consistent radiocarbon dates of human occupation no earlier than 1000 CE and more likely to 1200 CE. Exactly the time the palm forests were depleted, the stone statues (*moai*) were being erected, and environmentally the Polynesian rats were contributing to environmental imbalance. Reevaluation of Polynesian settlement of Hawaii and New Zealand dates also places their first human settlements to these times. Thus we might now see purposeful Polynesian expeditions all across the eastern Pacific for purposes of migration, settlement, and trade at much later dates than in the traditional interpretation. This is important and it may be related to the second find.

The second archaeological find of relevance involves Polynesian chickens in Chilean archaeological sediments dated between 1300 and 1420 CE. The dates fit well with later settlement of Easter Island and the eastern Pacific in general. Although debated, the published finds of New Zealand researchers Alice Storey, Elizabeth Matisoo-Smith, et al. (2007) at the El Arenal site in Chile appear legitimate. These dates are well before the Pizarro's expedition that conquered the Inca Empire and sent raiding parties down to Chile in the 1530s CE period. The expedition noted the presence of chickens in the Andes, but we cannot be sure of this description as camelids like the llamas were described as sheep. In any case, the Spanish introduced chickens into what became the viceroyalty of Peru, but their presence in modern Chile was quite light until later times. So history can shed little light on the presence or absence of prehistoric chickens in the Andes. However, archaeology can. Since no other domesticated chicken or fowl from pre-Columbian archaeological contexts have been found to date, we can suggest for now that the Chilean finds might well stem from Polynesian voyages that continued to the Americas in an easterly direction as they settled the eastern Pacific Ocean after 1200 CE.

Storey, Matisoo-Smith, et al.'s claims, of course, were controversial and the implications far reaching. A critique of those claims by Gongora et al. in 2008 was published in the same journal as the original claims. Storey, Matisoo-Smith, et al. reexamined their work and produced even stronger evidence for Polynesian chickens in prehistoric South America at El Arenal in Chile. The site was definitely abandoned before European contact and had no European artifacts. Two more chicken remains were uncovered from El Arenal and further support the original DNA analysis of these chickens as Polynesian. The dating of these remains to the late prehistoric or late pre-Columbian period (late 14th century CE median date) was not contaminated by marine deposits or food supply for the chickens. Their diet was terrestrial, not maritime. Further, Storey, et al. asserted that Cooper's study showed the drawback of using only modern DNA for ancient relationships when ancient DNA was also available. This refutal is probably also relevant for Easter Islander genetics studies as well. For the moment, the idea of Polynesian chickens in a pre-Columbian site in Chile seems well supported.

Our domesticated chickens also fit well with another domesticated food mystery, one that Thor Heyerdahl used to support his claims of South Americans on Easter Island. Sweet potatoes were first domesticated in the Americas but eventually showed up even on Easter Island. Though their seeds can germinate in salt water and they are found in some Polynesian settlements in the western Pacific, it seems also possible that Polynesian seafarers who brought the chicken to South America could also bring back the sweet potatoes to Easter Island. Claims that the Humboldt current do not allow such return voyaging except in El Niño storm years are problematic, as it has not been demonstrated that today's climate and weather patterns are the same as those 800 years ago.

The background of domesticated chickens could also support ties between Easter Island and South America. The wild ancestors of chickens are thought to derive from several varieties in Southeast Asia. Chickens may have been domesticated several times in Asia, perhaps independently. However, the archaic breed known as the Green Junglefowl and the Araucana chicken of Chile as well as Easter Island chickens show some affinities. Often tailless and laying blue-colored eggs, ties between these chickens suggest, along with giant palm trees and sweet potatoes, ties to Easter Island. But again, due to the nature of historic contact, these issues are presently difficult to explore further.

We would tend to expect chickens to come from Easter Island, as there were apparently no chickens in the pre-Columbian world before the plausible Polynesian contacts in South America, and chickens were the only major food supply animals brought by the early expedition(s) to settle Easter Island. No pigs or dogs have been found in a prehistoric context on Easter Island, and only the Polynesian rat may have been an additional food supply animal. This differs from other Polynesian settlement expeditions in the Pacific where pigs were a

common food supply, such as in the Hawaiian Island settlement expeditions. Chicken pens in archaeological contexts can be found all over Easter Island and were noted by European explorers including the 1774 CE Captain James Cook expedition. Assuming that no other Polynesian domesticates show up in South America, Easter Island is the closest and most likely source for Polynesian chickens arriving at El Arenal in Chile.

Finally, increased research in the past two decades in South America points to a much richer seafaring tradition than was previously assigned to pre-Columbian South Americans. In Inca times, that highland civilization, with little seafaring background, helped sponsor expeditions that led to Inca pottery on the Galapagos Islands. The Inca relied upon a 2,000-year-old seafaring tradition on the coast not unlike the Persian relationship with the Phoenicians in the Mediterranean Sea. The 2,000-year-old Moche civilization was known to bring in whales on their watercraft, but it is the later coastal civilizations who would have most likely interacted with incoming Polynesians, groups like the Inca in the south or the Chimu in the north of Peru. Several origin myths even indicate the founding of these civilizations by legendary seafaring leaders who came by boat, including Taycanamu, Naymlap, and Chimo Capac. Taycanamu, according to Spanish chronicles, arrived by sea with a large fleet, established the Chimu dynasty, and left his son in charge before heading back across the sea to the west. Even the legendary god figure Viracocha is said to have left the Andes by going across the sea.

Such myths are similar to the founding myths associated with Easter Island (also known as *Te Piti* or navel of the world). These suggest the founder Ariki Hotu Matua sent out an exploratory canoe expedition before arriving on the island in his own expedition at Anakena beach after leaving "Hiva" possibly from the Marquesas and perhaps Mangareva to the west. The expedition was well planned with various flora and fauna brought along for settlement and with multiple waves of settlers to increase the chances of success.

It was into this rich seafaring trade zone of the eastern Pacific that Polynesian traders would have arrived in the late pre-Columbian times in South America. Trade took South American metallurgical items as far away as west Mexico and had Spanish chroniclers describing merchant seafaring communities with 6,000 members and craft that held 30 people and utilized triangular sails. If indeed these civilizations encountered each other, then there are many possibilities concerning flow of people, direction of trade, and more that need to be reexamined.

Conclusion

The traditional interpretation of Easter Island human history, while based on high-quality research, like all research into a distant past where most of the

evidence has disappeared to the ravages of time, is in constant need of reevaluation and reinterpretation. This was true for Thor Heyerdahl's ideas, and it is true for today's research. New views that push the settlement of eastern Polynesia forward, especially Easter Island, New Zealand, and the Hawaiian Islands to around 1200 CE, challenge the traditional, established chronology and history for those islands and fits well with possible Polynesian contacts in the Americas after that time. These adjusted chronologies, historical descriptions of a still fertile Easter Island in the 1700s, and the role of the Polynesian rat, both as ravager of islands and as food source, also challenge traditional interpretations of the collapse of Easter Island. An Easter Island civilization that emerges later and survives longer than the traditional interpretations suggests is more likely to carry out expeditions for trade and contact with South America and elsewhere, contact that likely led to Polynesian chickens in pre-Columbian El Arenal in Chile. Chickens are exactly what we would expect to have come from Easter Island, as there were apparently no chickens in the pre-Columbian world before the plausible Polynesian contacts in South America, and chickens were the only major food supply animal brought by the early expedition(s) to settle Easter Island. This differs from other Polynesian settlement expeditions in the eastern Pacific where pigs were a common food supply, and this makes Easter Island by far the closest and most likely source for those chickens.

Such contact would also explain how South American-descent sweet potatoes show up on Easter Island. Dare we bring up the taboo subject of contact in both directions? Polynesian seafarers could have navigated their way back to Easter Island, especially in El Niño years, bringing South American foodstuffs such as sweet potatoes and even possibly, pre-Columbian South Americans. The research of Marshall Weisler (1995) on the Pitcairn Islands supports such an exchange model used by Polynesians to settle and create exchange systems on new islands. Resources and people were moving between the Pitcairns (Henderson and Pitcairn the nearest neighbors to Easter Island) and the Marquesas Islands, and such a model might well have connected Easter Island and South America. Because we can never genetically test all ancient populations on Easter Island or anywhere else, this can only be a suggestion. In other words, even if it is only Polynesians that settled Easter Island as part of an explosive push across the eastern Pacific, this could be seen as part of a larger interaction network, bringing up the question of trade and exchange with Easter Island's nearest large geographic resource, South America. The El Arenal chickens are strong evidence for something more than accidental contact, and the question can now be asked, what was the nature and extent of that contact?

Admittedly, we are dealing with a few chicken bones, hypothetical seafaring activities, and interpretation of data toward a set of historical events we may never be able to fully understand. We might ask ourselves—as researchers and

interested parties on the subjects of Easter Island, contact, and issues of diffusion and exchange—if similar research scenarios existed in the past. What was the outcome of continued research into such controversial areas? We need look no further than the case of Scandinavian seafaring Vikings and contact with North America in the pre-Columbian period. Historic Viking sagas known as the "Icelandic Sagas" suggested contact and settlement might have taken place in areas referred to as Greenland, Markland, and Vinland near North America. However, the normative interpretation was that the Vikings went no farther than Iceland in the North Atlantic. Then came the 19th-century findings of significant Viking settlements on Greenland where 3,000 to 5,000 Vikings settled with cattle and more on over 400 farms during the period of 980–1430 CE. Complex trade with both Europe and the native peoples of North America was a hallmark of Greenland. Expeditions to explore the mainland of North America from nearby Greenland took place as early as 985 CE. Even Vinland turned out to be real as 1960s excavations at an actual Viking longhouse and settlement located at L'Anse aux Meadows in Newfoundland proved. The settlement may not have been the only one, as other contacts and expeditions from Greenland are mentioned, but this one is archaeologically verifiable and dates to around 1000 CE. Will the chicken bones of El Arenal turn out to be part of a much larger story of contact and exchange, two classic hallmarks of human civilization, as was the case with Native Americans, Vikings, and North America? With only oral traditions, historic accounts, and archaeology to guide us in South America, it may not be possible to elucidate the Polynesian and South American contact, as was the case with the Vikings and Native Americans. New finds are always just over the next wave of evidence interpretation.

References and Further Reading

Barnes, S. S., Elizabeth Matisoo-Smith, and Terry Hunt. "Ancient DNA of the Pacific rat (Rattus exulans) from Rapa Nui (Easter Island)" *Journal of Archaeological Science* 33 (2006):1536–1540.

Barthel, Thomas. *The Eighth Land: The Polynesian Settlement of Easter Island*. Honolulu: University of Hawaii Press, 1978.

Cavilli-Sforza, Luigi, P. Menozzi, and A. Piazza. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press, 1993.

Diamond, Jerod. *Collapse: How Societies Choose to Fail or Succeed*. London: Penguin, 2005.

Dransfield, John. Palm and Cycad Societies of Australia. "Palms: *Paschalococos dispertab*." Available at www.pacsoa.org.au/palms/Paschalococos/disperta.html (accessed June 4, 2010).

Gongora, Jaime, et al. "Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA," *Proceedings of the National Academy of Sciences* 105:30 (July 29, 2008), 10308–10313.

Hagelberg, Erika. "DNA from Ancient Easter Islanders." *Nature* 369 (1994): 25–27.

Heydon, Doris, and Paul Gendrop. *Pre-Columbian Architecture of Mesoamerica*. New York: Rizzoli Publications, 1988.

Hunt, Terry L. "Rethinking Easter Island's Ecological Catastrophe." *Journal of Archaeological Science* 34 (2007): 485–502.

Matisoo-Smith, Elizabeth, and J. H. Robins. "Origins and Dispersals of Pacific Peoples: Evidence from mtDNA Phylogenies of the Pacific Rat." *Proceedings of the National Academy of Sciences* 101 (2004): 9167–72.

Matisoo-Smith, Elizabeth, R. M. Roberts, G. J. Irwin, J. S. Allen, D. Penny, and D. M Lambert. "Patterns of Prehistoric Human Mobility in Polynesia Indicated by mtDNA from the Pacific Rat." *Proceedings of the National Academy of Sciences* 95 (1998): 15145–50.

Millon, Rene. *Teotihuacan: Onceava Mesa Redonda*, Vol. I. Mexico City: Sociedad Mexicana de Antropologia, 1966.

Orliac, C. "The Woody Vegetation of Easter Island between the Early 14th and the mid-17th Centuries A.D." In *Easter Island Archaeology: Research on Early Rapanui Culture*, edited by C. Stevenson and W. Ayres. Los Osos, CA: Easter Island Foundation, 2000.

Rainbird, P. "A Message for Our Future? The Rapa Nui (Easter Island) Eco-disaster and Pacific Island Environments." *World Archaeology* 33 (2002): 436–51.

Storey, Alice, J. M. Ramirez, D. Quiroz, D. V. Burley, D. J. Addison, R. Walter, A. J. Anderson, T. L. Hunt, J. S. Athens, L. Huynen, and E. A. Matisoo-Smith. "Radiocarbon and DNA Evidence for a Pre-Columbian Introduction of Polynesian Chickens to Chile." *Proceedings of the National Academy of Sciences* 104, 25 (2007): 10335–39.

Urton, Gary. "Inca Myths." In *World of Myths*, Vol. II, edited by Felip Fernandez Armesto. Austin: University of Texas Press, 2004.

Weisler, Marshall. "Henderson Island Prehistory: Colonization and Extinction on a Remote Polynesian Island." In *The Pitcairn Islands: Biogeography, Ecology, and Prehistory*, edited by T. Benton and T. Spencer. London: Academic Press, 1995.

White, Nancy. "Late Intermediate Peru: Chimu and Other States A.D. 900–1400." Available at http://www.indiana.edu/~arch/saa/matrix/saa/saa_mod08.html (accessed August 3, 2010).

Early American Impressions of Rapa Nui

The following account, written by U.S. Navy Paymaster William J. Thompson, was included in the 1889 Annual Report of the Smithsonian U.S. National Museum, and presents some of the earliest reports of European and American reaction to the native peoples of Rapa Nui.

The native character and disposition has naturally improved as compared with the accounts given by the early navigators. They were then savages wearing no clothes, but with bodies painted in bright colors. The women are said to have been the most bold and licentious in Polynesia, if the reports are correctly stated, but we found them modest and retiring and of higher moral character than any of the islanders. The repulsive habit of piercing the lobe of the ear and distending the hole until it could contain bone, or wooden ornaments of great size is no longer practiced, but there are still on the island persons with earlobes so long that they hang pendent upon the shoulder. In disposition the natives are cheerful and contented. Our guides were continually joking with each other, and we saw no quarreling or fighting. They are said to be brave and fearless of danger, but revengeful and savage when aroused. They are fond of dress and ornaments. Very little tappa cloth is now worn, the people being pretty well equipped with more comfortable garments, obtained from the vessels that have called at the island. . . . Straw hats are neatly braided by the women and worn by both sexes. The women wear the hair in long plaits down the back, the men cut the hair short and never discolor it with time as is the custom in many of the islands of Polynesia. The hair is coarse, black, and straight, sometimes wavy, but never in the kinky stage. The beard is thin and sparse. Gray hair is common among those beyond middle life and baldness is very rare.

Source: William J. Thomson. *Te Pito te Henua; or, Easter Island*. Washington DC: Government Printing Office, 1891.

CON

Thor Heyerdahl's expedition that he and his crew undertook with the balsa raft *Kon-Tiki* (an old name for the pre-Columbian Andean sun god Viracocha) in 1947 was not the first voyage with a traditional vessel in the Pacific Ocean. Heyerdahl had been inspired and challenged by the seafaring explorations carried out by the Frenchman Eric de Bisschop in the 1930s. De Bisschop's first voyage in a Chinese junk took him from Taiwan to Australia. Another, more spectacular voyage was undertaken shortly before the outbreak of World War II, this time from Hawaii to France in a Polynesian double-hulled vessel that de Bisschop named *Kaimiloa*.

While de Bisschop's intention was to demonstrate that seafaring in a west-east direction into Polynesia and within its vast expanse were possible in

prehistoric times, Heyerdahl set out to prove the possibility of east-west seafaring (i.e., from South America into Polynesia). The *Kon-Tiki* expedition was repeated almost 60 years later, in 2006, when Torgeir S. Higraff set out with his crew in a balsa raft called *Tangaroa* (which is the name of the Maori "God of the Seas"). The *Tangaroa* expedition was intended predominantly to carry out investigations of the ecological conditions of the Pacific fauna.

The *Kon-Tiki* expedition in turn challenged de Bisschop to navigate a route from west to east (from Tahiti to South America) in a bamboo raft named *Tahiti Nui* (Great Tahiti). It is noteworthy that both Heyerdahl (in 1947) and De Bisschoff (in 1957) lost their vessels because they shipwrecked, *Kon-Tiki* in the Tuamotu Archipelago and *Tangaroa* near the Juan Fernandez Islands off the coast of Chile. But both voyages proved that seafaring in traditional vessels was possible in either direction.

Although the *Kon-Tiki* endeavor stirred up much debate about American-Polynesian contacts in prehistory, it is De Bischoff's merit to have directed attention to a topic that was gaining in momentum, namely the investigation of possible landings of Polynesian seafarers on the shores of South America. The recent find of a bone of a Polynesian chicken in Chile may provide a lead for such investigation.

The *Kon-Tiki* expedition has remained the most spectacular of the Pacific explorations in rebuilt traditional vessels. This is also true for the conjectures and speculations about seafaring in the Pacific and about the peopling of the Polynesian islands that Heyerdahl publicized with his famous book *Kon-Tiki* (1950). Half a century ago, it may have still been possible to take Heyerdahl's claim about a migration from South America to Easter Island at face value. Then, the *Kon-Tiki* expedition could be categorized as an exercise in applied oceanographic science, with the weight to mark, in reality, the sea route where theory had it. However, nowadays, after decades of research in the prehistory and human genetics of Polynesian populations, Heyerdahl's enterprise has lost its validity as proof for prehistoric voyages from America to the West. In the history of Polynesian studies, the *Kon-Tiki* expedition will always retain its value as a demonstration of what is technically possible, although the insights from modern research do not back up Heyerdahl's claims.

Heyerdahl had tried to substantiate his claim about an alleged American population transfer to Polynesia with a serological study, comparing data from Peruvians and Polynesians. The invalidity of that study was convincingly proven by R. C. Suggs already in 1960. In view of the heated debate that arose in American academic circles in the 1950s and early 1960s, one cannot but admit "Heyerdahl's role as a pleasant *advocatus diaboli* who has provoked a mighty upsurge of archeological field-work in the Pacific area" (Barthel 1963: 422).

In 1956, Heyerdahl impressed the public with another exercise of applied science, this time with his demonstration how the *moai*, the large memorial

stones on Easter Island, could have been erected. With traditional technical equipment, Heyerdahl and his crew succeeded in hauling several of the stones that had been lying on the ground, into an upright position, thus restoring a state of the cultural space prior to the arrival of Europeans.

What was demonstrated by this spectacular action were the possibilities of traditional engineering. Heyerdahl presented the world with a technical demonstration of how the stones could have been transported and hauled. In his book *Aku-Aku* (1958), the endeavor of 1956 is documented and recorded. From the standpoint of history, however, Heyerdahl's reconstructions fall short of a proper cultural embedding. Heyerdahl's technical action is no proof that the islanders did so in historical times. Modern anthropological research has produced insights that speak in favor of the *moai* as having been transported in an upright position straight from the quarry on the slope of Rano Raraku, one of the volcanoes, down to the lowland.

Kon-Tiki and *Aku-Aku* were illustrative demonstrations of modern Euro-American problem solving. In both enterprises, technical possibilities were explored to the fullest by the determined mind, without properly relating them to the conditions set by local cultural history. These conditions can be reconstructed to form an overall picture of Easter Island and the culture of its inhabitants, which is Polynesian to the core. Easter Island is called Rapa Nui (Flat High Plateau) by the islanders, and their Polynesian ties will be highlighted in the following for several domains, reconciling data from human genetics, biology, linguistics, cultural anthropology and the study of mythology, and popular beliefs.

Polynesian relationships are evident on Easter Island, not only with respect to human ecology (i.e., colonization, cultural history, and language) but also concerning environmental ecology (i.e., transfer of plant and animal species). Within the Polynesian network, the culture of Rapa Nui is most closely affiliated to that of the Marquesas Islands, at a distance of some 3,400 km to the northwest.

The Polynesian Genetic "Footprint" on Easter Island

On the genetic map of human populations in the Pacific region, Easter Island is the easternmost outlet in a wider, uniform landscape of a Polynesian gene pool. Although the dynamism of genetic expansion throughout the Pacific is still a matter of debate, there is consensus that the genomic profile of the Rapa Nui islanders is affiliated to the Polynesian ethnic stock.

Together with the Society Islands (including Tahiti), the Cook Islands, and New Zealand, the Rapa Nui genome forms part of "a cluster that is at a considerable distance from the rest of the Pacific islands" (Cavalli-Sforza et al. 1994: 364). The position of Easter Island as a genetic outlier conforms with the history of Polynesian migrations. Easter Island was reached by migrants coming from the northwest at the beginning of the fourth migration wave in the Pacific region (CA. 300 CE).

If there had been a pre-Polynesian population of American descent on Easter Island, then one would expect traces of ancient genetic mixing in the islanders' gene pool. The genetic admixture that is found in the genome of some islanders dates to the 19th and 20th centuries and is due to their descent from ethnically mixed marriages (i.e., of Rapa Nui people with Chilean immigrants to the island).

Plant and Animal Import from Polynesia to Easter Island

The history of human settlement on Easter Island is the history of the genetic drift from the center to the periphery of Polynesia. The same west–east drift can be established for the species of nonindigenous flora and fauna that were imported to the island by the early settlers. The origin of all those species has yet to be found in the ecological environment of Polynesia.

- Imported plants (with their names in Rapa Nui and in Latin for botanical identification):

    Kumara: various species of batatas (sweet potatoes; *Ipomoea batata*);

    Taro: the roots of which are edible (*Colocasia esculenta*);

    Uhi: various species of yams (*Dioscorea* sp.);

    Kaha: pumpkin (*Lagenaria siceraria*);

    Maika: banana (*Musa sapientum*);

    Toa: sugarcane (*Saccharum officinarum*);

    Ti: species of agave (*Cordyline terminalis*);

    Pua: Curcuma/yellow root (*Curcuma longa*);

    Mahute: species of mulberry (*Broussonetia papyrifera*) whose bark were worked to obtain fibers for making textiles (*tapa*);

    Marikuru: a tree species with fruits containing a soap-like syrup (*Sapindus saponaria*);

    Mako'i: a tree species whose wood was used for carving (*Thespesia populnea*);

    Naunau: a species of santel wood.
- Imported animals: The only two animal species that were not indigenous and came with the migrants were the Polynesian chicken (*moa*) and—unintentionally imported—the Polynesian rat (kio'e—*Rattus concolor*).

The history of spread of these various species provides evidence for Polynesian origins of the settlers and stands against speculations of an American colonization of Easter Island. Imports from America date to modern times. For instance, in the caves on the island, the poisonous spider, whose popular name

is "black widow," finds its habitat. It is assumed that this spider species reached Easter Island in the 20th century with ship cargo from Chile.

Rapa Nui as a Polynesian Language

The indigenous language spoken by the islanders, Rapa Nui, is affiliated with the Oceanic subgroup of Malayo-Polynesian languages which are a branch of the Austronesian phylum. Rapa Nui shares all the major features of its linguistic structures with the other Oceanic languages. The sound system is extremely simple. It consists of five vowels and nine consonants (as opposed to eight in Tahitian or Hawaiian). A glottal stop may separate vowels, and its occurrence causes semantic change, that is words with an identical sound sequence have different meanings in case the glottal stop occurs: compare *pua* "flower": *pùa* "to cover oneself"; *hau* "string": *hàu* "hat." Word formation is limited to the use of some prefixes and suffixes. The grammatical structure is predominantly analytical. For instance, to mark the cases in noun inflection, prepositions are used that precede the noun (*o tou* for the genitive, *i tou* for the accusative); for example, *o tou poki* "of the boy."

The vocabulary of Rapa Nui shares, with other Oceanic languages, basic terminology in various domains such as seafaring (vessels and sailing), pottery for cooking and storage, horticulture (e.g., banana, yams), domesticated animals (i.e., fowl), constructions (e.g., housepost), weaving (e.g., weaving spindle), religion, and ceremonial traditions. Modern language use includes numerous lexical borrowings from Spanish and various English internationalisms. Bilingualism of the islanders (with Rapa Nui as first, and Spanish as second language) has produced patterns of interference of the linguistic systems, including code-switching.

If Heyerdahl's claim of American origins had any factual background, then the language of Easter Island would be somehow affiliated to the linguistic landscape of South America. In other words, if the islanders' ancestors had been migrants from America, they would have brought their language with them to Easter Island. Regarding any kind of relationship with American languages, the record for Rapa Nui is negative, no matter how "linguistic affiliation" is defined:

- As a member of the Oceanic group of the Malayo-Polynesian branch of the Austronesian phylum, Rapa Nui is genealogically distinct (and separated) from languages in the Americas, and from those in South America in particular.
- Rapa Nui is a genuinely Polynesian language without any admixture of an American language. This would be recognizable in its sound system, its grammatical structures, and its vocabulary. Such admixture would point in the direction of a fusion (or mixed) language, which Rapa Nui is not.
- In Rapa Nui, no ancient relics of a non-Oceanic language can be identified that would be evidence for a pre-Polynesian speech community. No language

vanishes from the record without leaving traces in the region where it was spoken, regardless of how many other languages may form an overlay. If there had been any American language spoken on the island, its prehistoric presence would be recognizable in the form of a substratum (a Latin compound word: sub- "under" + stratum "layer"), that is, as elements of an underlying residue.

- The consequences of intense language contact have been studied for many settings throughout the world. The findings from extensive comparative research show that the interference of a language (Language A) that is over-formed by another (Language B) may continue for a prolonged period, even if the overformed languages becomes extinct. This is a reflection of a stage of bilingualism when speakers of Language A transfer certain habits (e.g., of pronunciation or phraseology) to Language B to which they eventually shift. An illustrative example of such conditions is the prolonged influence of Etruscan (the pre-Roman language of Tuscany in Italy) on Latin in antiquity and on the local Italian dialect (i.e., Tuscan). The aspiration of certain consonants in Tuscan is a regular phenomenon in the habit of pronouncing sounds, which has been perpetuated over many generations and is explained as a long-term influence from Etruscan, still recognizable 2,000 years after Etruscan became extinct.

- The relevance of substratum influence is true also for naming. Although on a relatively small scale, there is much variation of the landscape on Easter Island. Natural formations include prominent markers such as volcanoes or high cliffs, in addition to plains and shallow beaches, crater ponds, springs and caves, and some offshore islets. All these parts of the landscape were integrated into the cultural ecology of the islanders through a process of name-giving, and the linguistic structures of those names are purely Polynesian. On Easter Island, no traces of pre-Polynesian names can be identified that would reveal any kind of linguistic affiliation with Amerindian languages.

- The perseverance of older names is a typical phenomenon of substratum continuity in all parts of the world. For example, there are many names of places, rivers, mountains, and of entire regions of Native American origin in the United States that bear witness to the former presence of Amerindians (e.g., Manhattan, Mississippi, Massachusetts). If there had been any Amerindian colonization of Easter Island, a distant memory of this would be reflected in name-giving. However, no such linguistic reflection exists.

- The maintenance of Rapa Nui is amazing since this language had been on the verge of extinction in the 19th century when only a few hundred indigenous inhabitants of Easter Island were left. Thanks to the resilience of the islanders, Rapa Nui has survived and is today spoken by some 3,000 people, of whom the majority live on the island. Some outgroups of Rapa Nui speakers are found in Chile and Peru.

The Cultural Heritage of the Rapa Nui Islanders

In addition to the local language, there is the web of Polynesian traditions that make up the cultural heritage of the islanders. The most visible of these traditions are the *moai* ("statue holding the spirit of an ancestor"), and they, literally, provide "massive" evidence for cultural ties with other regions of Polynesia. The *moai* were erected on ceremonial platforms (called *ahu* in Rapa Nui) near the beach and they faced inland, "looking over" the village community. Those *moai* that can still be seen today in an upright position in their original space were all reerected in the course of restaturation of the island's cultural heritage.

Some 300 ceremonial platforms have been identified, and there may have been more than 1,500 *moai* standing at sites scattered throughout the island. In the main quarry, on the slopes of Rano Raraku, where the lava blocks were obtained from which the statues were carved, as many as 400 *moai* in various stages of their carving can be seen. The average height of the *moai* is 4 meters and the average weight is about 50 tons. The tallest statue that was ever erected on an *ahu* is 10 meters high and weighs 80 tons. It stood on the Ahu Te Pito te Kura in the northeast. The biggest and heaviest *moai* that was ever carved on Easter Island remained unfinished and is still fixed to the rock in the quarry. Its length is 22 meters and its weight has been estimated at 250 tons. Some *moai* were adorned with an additional (separate) headpiece (*pukao*) which most probably imitated the traditional headdress of the islanders.

The *moai* were the visible expression of a vivid ancestor cult among the islanders. The ceremonial reverence for the ancestors is an ancient custom widespread in Polynesian cultures. On Easter Island, *moai* were erected during a period that lasted for some 800 years, from CA. 700 CE to the beginning of the 16th century. During the long span of time, the outer appearance of the statues experienced changes, in the proportions of body parts, of facial features as well as regarding their stylistic realization.

The types of *moai* exhibit stylistic resemblances with statuary known from other parts of Polynesia. The most striking similarity lies with statues from the Marquesas Islands. Legend has it that the early migrants came from the direction
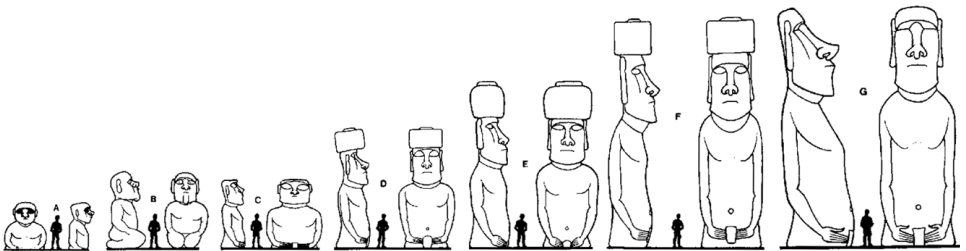
Typology of ancestor statues (*moai*) on Easter Island. (after Charola 1997: 60)

of the Marquesas (see below for mythology), which would explain how the artistic and aesthetic memory of the migrants' homeland (i.e., the Marquesas) was transmitted to and revived on Easter Island. The types of *moai* show different degrees of a local development of statuary, culminating in the most deviant types.

The cult of the *moai* experienced a decline in the early 16th century, as a consequence of the turmoil of intertribal clashes. It has been estimated that, at the dawn of prolonged conflicts among the clans over shortening natural resources, the population of the island may have counted between 7,000 and 10,000 individuals. The subsistence crisis put an end to activities relating to the carving and erecting of statues and to ceremonial services at the *ahu*. Possibly, the economic crisis triggered a crisis in the belief system. The revered ancestors, whose spirits had been watching over the daily affairs of the living members in the community for generations, seemed to be helpless in a time of conflict. Eventually the lack of trust in the protective capacities of the ancestor spirits may have led to the abandonment of the *moai* cult.

Community life on Easter Island continued, although it was overarched by a new worldview, the birdman cult (see the section below "Specifically Local Cultural Innovations on Easter Island").

Mythology and Popular Beliefs

From the beginnings of social relations among modern humans, oral tradition has been a prominent source for organizing the collective experience of the community in the cultural process. Arguably, the oldest text genre that developed was the explanation of the world within the framework of mythopoetic conceptions. "The myth is the prototypal, fundamental, integrative mind tool" (Donald 1991: 215). The most elementary layer of mythopoetic experience that we find in the world's cultures are myths of origin, usually explaining, in ethnocentric terms, how a certain group of people (a clan or kin group) is in the focus of historical events.

The myth of origin the inhabitants of Easter Island have created for themselves is a prominent marker where their cultural identity crystallizes. According to myth, the ancestors of the islanders came from the Marquesas Archipelago, in particular from Hiva Island. The legendary hero who led the colonizers to their destination was King Hotu Matu'a. In the mythical tradition, one single enterprise of colonization is recorded. The insights from modern anthropological and genetic research confirm that the settlement of the island was a onetime event.

After Hotu Matu'a and his followers had set foot on the Easter Island, the newcomers lived in isolation until the arrival of the Dutch captain Jacob Roggeveen, who discovered the island on Easter Sunday in 1722. The island, named Te Pito o te Henua ("the navel, center of the earth"), was divided by Hotu Matu'a into two parts, which were called Kote Mata Nui (in the north and west) and Kote Mata Iti (in the east and south).

The northern region was inhabited by the clans of Tu'uaro, the southern region by the Hotu'iti. The son and the nephew of Hotu Matu'a were revered as the founding ancestors of the clans. The internal structure of a clan was based on the units of *ure* (lineage) and *ivi* (extended family). The royal family (*ariki paka*) and the priesthood (*ivi atua*) represented the social elite.

The myths and popular beliefs among the islanders are permeated by Polynesian interconnections, and there are no clues as to early contacts with America. If there had been several waves of colonization, including one directed from South America, the memory of such events would have been crystallized in oral tradition and imagery. However, the Polynesian heritage is all there is in Rapa Nui mythology.

Specifically Local Cultural Innovations on Easter Island

It would be a misconception to perceive human ecology of Easter Island as a mere distant outlet of the Polynesian cultural network. In the course of time, Rapa Nui culture has produced its own local innovations, which are as significant for its fabric as are the traditional features the islanders share with other Polynesian societies.

The Birdman Cult

After the demise of the *moai* cult, a new worldview emerged amid continuous intertribal conflict. In the strive for a unifying symbol that could be shared by all islanders beyond clan rivalry the cult of the birdman was formed. This mythical being (Tangata Manu)—with a bird's head and human limbs—was identified as the representative on Earth of the newly established creator god Makemake. Images of the birdman are especially numerous in the area of Orongo on the slope of the volcano Rano Kau at the southwestern tip of the island. The motif of the birdman is mostly depicted in rock carvings (i.e., petroglyphs).

Orongo is a sacred site where an important annual ceremony was held. From a high cliff, with a wide view over the ocean, one can watch the

Motif of the birdman from Easter Island. (after Lee 1992: 1)

arrival of the sea birds in spring. Among them is the manu tara ("sooty tern" in two subspecies, *Sterna fuscata* and *Sterna lunata*) that nests on three small islets (i.e., Motu Kaokao, Motu Iti, and Motu Nui) off the southwestern coast of Easter Island. It was the eggs of this bird that played a central role in the spring ceremony.

The protagonists of this ceremony were birdman contestants, aides (*hopu*) to the chiefs of the clans. They had to swim—in shark-infested waters with strong currents—to cross the strait that separates the small islets from Easter Island. The purpose of the contest was to retrieve an egg from the nest of a sooty tern and transport it safely back to the main island. The most difficult part of the contest was the task of climbing up the steep slope of the cliff on whose top the chiefs of the clans had gathered. The swimmer who arrived first and gave the intact egg to his chief was the winner.

Winning the birdman contest was not only a matter of individual bravery but it also meant the concession of privileges to the clan of which the contestant was a member. The winner, as the personification of the birdman, had to lead a secluded life for the rest of the year and live in a cave or ritual house, near the ancient quarry of Rano Raraku or near the northern coast at Anakena, which is the place where Hotu Matu'a landed according to the myth of origin. For one year, the winning clan ruled over the island and took care of the islanders' affairs.

The birdman cult fell into decline after Christian missionaries exerted their influence to convert the indigenous population. And yet, Christianity in its Catholic version never succeeded in erasing the memory of the birdman and its popularity. Symbolic of this is the imagery in the main church on the island at Hangaroa. Near the entrance door stands a large figure of the birdman, who is integrated into the local Christian canon and venerated as a saint. The motif of the bird is also present in the Christian imagery itself. In the sculptures—of the Virgin Mary holding her child and of the adult Jesus—that serve as altar pieces, the bird crowns the headdress of the human figures.

Writing Technology

Among the innovations that Rapa Nui culture experienced in precolonial times is the elaboration of a script. Documentation of the Rapa Nui writing system is found on the *Rongorongo* ("wood tablets/boards with incised script signs"). The *Rongorongo* script is the most elaborate of all the systems of notation in Polynesia and the only form of systematic sign use that meets the requirements of writing technology. Of the Polynesian languages, only Rapa Nui was written. The other speech communities of Polynesia were illiterate before the advent of the Europeans. During colonial times, the Latin alphabet was introduced to write the major Polynesian languages, among them Tahitian, Fijian, Maori, Hawaiian, and others.

The Latin script was also adopted to write Rapa Nui in the 20th century. Then, the knowledge of how to write and read *Rongorongo* texts had already been lost. The keepers of this knowledge were the priests (*ivi atua*), who were called *taula* in other parts of Polynesia. In the 1860s, Easter Island was depopulated by raids of slave traders who deported the majority of the islanders to Peru to work as laborers in the mines. Among them were the last of the priests, and the secrets of *Rongorongo* writing vanished with them.

The *Rongorongo* tablets were kept in caves, at sacred sites. Many of these tablets got lost, they were burnt by Christian priests, or they decayed under the weather conditions. How many *Rongorongo* might have ended up in private collections will remain a secret. Less than 30 texts are known to have survived. In some of the legends of the islanders the *Rongorongo* are described as "talking boards." Most probably, the contents of the *Rongorongo* texts was ceremonial (containing recitations, prayers, ritual chants, incantations, and mythical narratives) to serve the priest in his role as master of ceremonies and keeper of sacred knowledge. The cultural embedding of the *Rongorongo* tradition points at it being a sacred script, but it did not serve the purposes of practical writing in daily affairs.

Several attempts have been made to decipher the script. Although progress has been made with the identification of the principle of writing and with the identification of individual signs, it is not yet possible to read or translate entire text parts. There is consensus that the *Rongorongo* script is logographic, that is, based on whole-word writing. This means that one sign stands for an entire word (or idea, respectively). As for the association of the sign system with the sounds of the Rapa Nui language, this may not have been realized at all or, if so, on a minimal scale. *Rongorongo* signs, therefore, may stand for Rapa Nui words of different length (i.e., from one to four or even five syllables) without rendering individual sounds as phonetic units. This kind of writing was typical
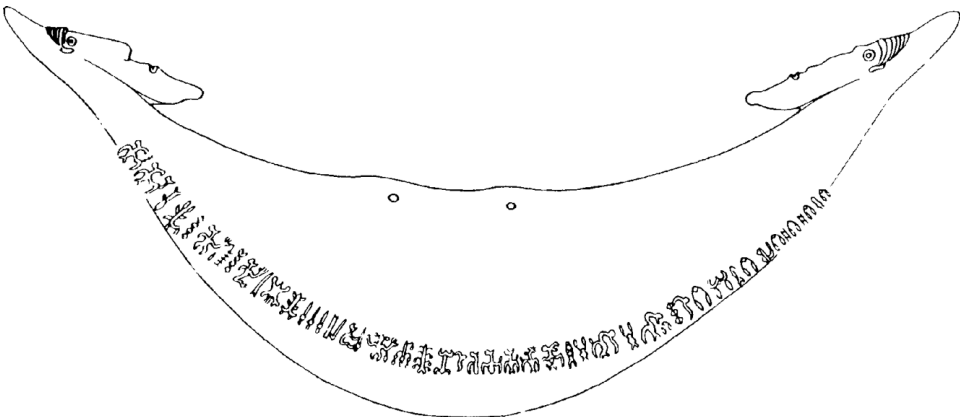
Rongorongo text from Easter Island (inscription as an adornment on a ceremonial pectoral, or *rei miro*). (after Haarmann 2000: 204)

of the early stages of writing technology, in the Old World (i.e., Danube script, ancient Sumerian pictography, ancient Indus script, Chinese writing of the oracle bone inscriptions) and in pre-Columbian America (i.e., Olmec writing).

The sign inventory is comprised of several hundred units. The visual forms of the signs show clear resemblances with the motifs in the rock art of Easter Island (e.g., images of the birdman, of vulvas, of pectorals, of geometrical motifs). The visual impression of the signs and the text corpus leave no doubt that *Rongorongo* was a genuine innovation of Rapa Nui coinage. When we ask for the motivation to elaborate such a script, then we find clues in an ancient Polynesian memorizing technique.

> Sacred chants, incantations, and genealogies were important throughout Polynesia, and these had to be recited exactly in order for the desired result to take place. Accordingly, priest-chanters on Easter Island used rongorongo tablets as mnemonic devices. These tablets probably were related to the chanting staves used in Mangareva and the Marquesas. (Lee 1992: 126)

References and Further Reading

Barthel, Thomas S., Review of *Archaeology of Easter Island*. Edited by Thor Heyerdahl and Edwin N. Ferdon, Jr., *American Anthropologist* 65:2 (1963), 421–425.

Cavalli-Sforza, L., P. Menozzi, and A. Piazza. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press, 1994.

Charola, A. E. *Isla de Pascua. El patrimonio y su conservación*. New York: World Monuments Fund, 1997.

Donald, M. Origins of the Modern Mind. *Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press, 1991.

Ferdon, E. N. *Science* 142 (December 1963):.

Fischer, S. R. *Rongorongo: The Easter Island Script—History, Traditions, Texts*. Oxford: Oxford University Press, 1997.

Du Feu, V. *Rapa Nui (Descriptive Grammar)*. London: Routledge, 1995.

Haarmann, Harald. *Historia universal de la escritura*. Madrid: Gredos, 2000.

Heyerdahl, Thor. *Aku-Aku: The Secret of Easter Island*. London: Allen & Unwin, 1958.

Heyerdahl, Thor. *Kon-Tiki Expedition: By Raft across the South Seas*. Translated by F.H Lyon. London: Allen & Unwin, 1950.

Langdon, R., and D. Tryon. *Language of Easter Island. Its Development and Eastern Polynesian Relationship*. Laie, HI: Institute for Polynesian Studies, 1983.

Lee, G. *The Rock Art of Easter Island. Symbols of Power, Prayers to the Gods*. Berkeley: University of California Press, 1992.

Pawley, A. The Austronesian Dispersal: Languages, Technologies and People. In *Examining the Farming/Language Dispersal Hypothesis* (pp. 251–73). Edited by P. Bellwood and C. Renfrew. Cambridge, UK: McDonald Institute for Archaeological Research, 2002.

Suggs, R. C. *The Island Civilizations of Polynesia*. New York: New American Library, 1960.

Van Tilburg, J. A., and J. Mack. *Easter Island: Archaeology, Ecology, and Culture*. Washington DC: Smithsonian Institution Press, 1995.

11

The Roman Empire's collapse was primarily due to social and political problems rather than the Barbarian invasions.

PRO	Heather Buchanan
CON	László Kocsis

PRO

Arguably the greatest empire of the ancient world, the Roman Empire eventually declined and collapsed. Exactly how the end came is still a matter of great debate today. At its height, the Roman Empire's reach was vast, extending from the Atlantic Ocean to the Euphrates River and from Britain to the Sahara. One thing appears clear: the barbarian invasions from without that ensued were the final blow to an empire already in the throes of collapse. At the core of the decline was the fact that the ideal of democracy was never truly realized in the Roman Empire; power always lay in either the hands of the few (as with the aristocrat-dominated senate) or in the hands of an emperor with absolute authority. Depending on the strength or weakness of a given emperor, so too went the internal stability of the empire. From a modern vantage point, Rome teaches how government can succeed and then fail due to the gradual internal crises left unchecked. The ensuing bureaucratic, political, social, and economic problems from the early days of the empire grew too numerous and too complex to be resolved in the end. Class conflict, constitutional flaws, and endless wars of conquest all contributed to the demise of the Roman Empire. Long before barbarian invasions came from without, Rome was collapsing from within.

Class Conflict

Stratification of Roman Society

Rome's first political and social steps after its transformation into a republic were problematic from the outset. Rome's humble beginnings was as a small Italian city–state consisting of an early Indo-European monarchy as the form of government until the early fifth century BCE. Around the sixth century BCE, the kings of the city–state were overthrown, and an aristocratic government was set up in place of the monarchy. In a true republic, power is concentrated in the

229

hands of the people, but in the case of the Roman republic, however, the power was concentrated within a small landholding nobility. In turn, these aristocratic families, the *patricians*, set themselves apart from the commoners, the *plebeians*. The patricians were primarily the members of the senate. After the establishment of the early republic, Rome became mired in its own bureaucratic machine.

The division of Roman society guaranteed that the mass poor remained at the bottom ranks. The division between the patricians and plebeians (who outnumbered the patricians) would serve to hurt the empire from within, as the plebeians were determined to gain political social equality. A military crisis with outside invaders around 450 BCE forced the patricians to ask for the plebeians' help, and only then were important concessions made in favor of the plebeians. A new Roman army was formed comprising plebeians, and soldiers were grouped into centuries, or hundreds, which created a new lawmaking body, the *comitia centuriata* (Assembly of the Centuries) (Jones 1989: 215–19). While the *centuriata* gained voting rights in the Roman republic, the resulting laws and officials needed to manage the new government brought problems. The growth of administrative business within the government required the creation of additional offices, such as *quaestors* (tax collectors and army administrators) and censors (assigned citizens to the new classes and the determination of senate membership). The first codification of Roman law, known as the Law of the Twelve Tables, became binding over all Roman people, aristocrats and commoners alike.

By 265 BCE (after its emergence from Etruscan dominance and an early sack by Gauls in 390 BCE), Rome dominated the Italian peninsula. The government at that time consisted of two elected patrician magistrates who had military and administrative powers, including oversight of the senate. This monopoly of the government by the patricians caused a rift with the plebeians, who threatened to secede from the government.

An unequal division of classes continued into the third century. The Punic Wars (264–149 BCE) led to the formation of new political factions in Rome: the senators, equestrians, and proletarians. The senators had great wealth but were prohibited by law to engage in trade or industry. These landholders tended to monopolize the senate. The equestrians were a monied class that had made its fortune in war contracts from the Punic Wars. The proletarians were primarily the landless poor and unskilled laborers, but they were citizens with voting privileges.

Originally, the republic consisted of small grain farmers; after the wars, however, these farmers were ruined by the influx of a million bushels of grain brought in from conquered Sicily in tribute. The grain, produced on large slave plantations, was sold in Rome at lower prices than that of the Italian farms. As a result, displaced farmers went to Rome in search of work. Their abandoned farms fell to the senatorial, aristocratic class; these landholders then invested in slaves and produced pricey products such as wine and olive oil, which were profitable until the later years of the empire.

Slavery

Slavery had always been a part of ancient civilization, with the leisure class forcing the multitudes to work their lands. Rome, becoming rich after the destruction of Carthage and Corinth in 146 BCE, took slavery to its extremes, with thousands of slaves bought and sold within the same day. The Romans had little regard for human life, and the exploitation of the peasantry was justified to sustain and support the landed class, which was regarded as the highest and noblest rank in society. As a result, cheap slave labor would lead to slave gang revolts, and by the time of Spartacus's revolts (73–71 BCE), the empire nearly fell. A large rural and urban slave population remained. The identity of what it meant to be "Roman" changed by the last two centuries of the empire, because emancipated slaves of Greek and Asian origin became Roman citizens.

Bureaucratic Corruption

The creation of a bureaucratic farm tax collection system led to corruption, which weakened the empire, for it corrupted officials and further demoralized the lower class. Between 265–133 BCE, Rome made imperial acquisitions outside of Italy: Spain, Africa, Macedonia, and Asia. This growing empire had to be administered, defended, and taxed. Each new province needed a Roman representative, a magistrate, to govern it. In the early provinces, Rome created
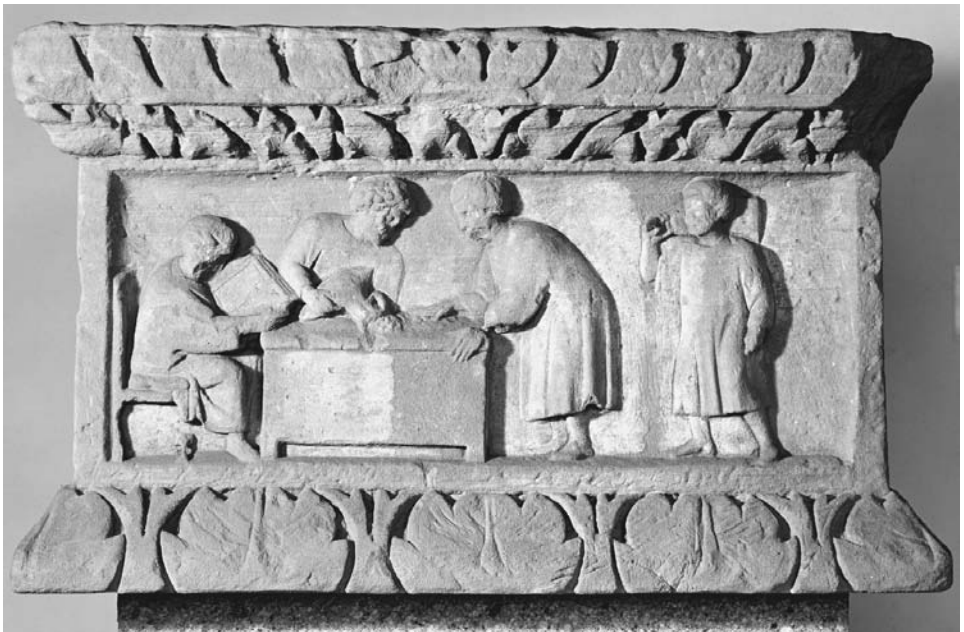
Relief depicting a tax collecting scene, Gallo-Roman, second century CE. (Rheinisches Landesmuseum, Trier, Germany/Giraudon/The Bridgeman Art Library)

praetorships with the power to rule. Each *praetor* was then given an assistant, a *quaestor*. Praetors made a fortune from their provincial commands: they accepted "gifts" from grateful subjects and made illegal assessments. A particularly lucrative method was to extend arbitrarily the limits of a given province through warfare; as the praetor was commander-in-chief of that region, he received a large share of the plunder from any and all subsequent conquests.

Like the praetors, tax collectors benefited from an inefficient and corrupt system. A curious feature of the Roman government was that tax collection was not one of the praetor's powers; instead, *quaestors* received the taxes from *publicani*, tax farmers. The *publicani* were Roman citizens organized into companies that bid for the right to collect taxes in a certain province for a set number of years. The system enabled the *publicani* to make a considerable profit above and beyond what was owed to the government. This system was without supervision, and the government appeared indifferent to the abuse of provincials by the tax collectors.

The growing class of businessmen took notice of the wealth to be gained and began taking their portion of the spoils of the empire. However, the rivalry between their greed with the public officials led to conflict. In Rome's capital, the poorer citizenry wanted some of the imperial wealth as well. Political machinations with the goal of imperial wealth added to the problems within the republic. Rome became the financial capital of the world. The new wealth, however, led to a Roman extravagance for the upper class that compromised the morals of a once dignified and rational society and left the landless poor with nothing.

Civil Unrest

Prior to the late second century BCE, disputes between the plebeians and patricians had always been resolved by negotiation and political reforms. The following century, however, saw partisan violence and civil wars that paved the way for the rise of military dictators such as Caesar, Antony, and Octavian.

Rome faced challenges from its landless citizens. Senators illegally claimed forfeited lands, and the proletarians were overrunning Rome itself. Senators were also creating scandals in the courts. Since all senators were jurors in court cases dealing with administrative scandals in the provinces, they rarely brought their fellow senators to trial for their crimes, particularly crimes they'd committed themselves. The Gracchus brothers, Tiberius and Gaius, came to political leadership in 133 and 123 BCE, respectively, and they tried to institute democratic reforms. In an attempt to help the proletarians gain land, Tiberius sponsored colonization programs that would distribute unassigned public lands in Italy to the landless citizens. However, senators fought this, and street riots broke out in the city. Tiberius himself was killed in street fighting. Ten years later, Gaius revived and expanded his brother's colonization program and went

so far as to form a coalition of proletarians and equestrians to combat the senators. Then, Gaius made the fatal move to extend citizenship to Italian allies. Unwilling to share citizen rights with anyone else, all three classes, the senators, equestrians, and proletarians turned against Gaius and he eventually killed himself after being hunted down.

By 90 BCE, Roman allies were demanding full citizenship or complete independence from the empire. Greek and Asian provinces were on the brink of revolt because of mistreatment and neglect of the Roman fleet, which was leading to piracy. A social war (90–88 BCE) took place between Rome and her allies. Rome was unprepared and had to quickly grant citizenship to stop the fighting. However, the citizenship was in name only, for the allies could not vote or take office. For 60 years after the social war, Rome was awash in riots, political assassinations, civil war, and general collapse in government. Crime became rampant, and criminal elements carried out bribery, blackmail, and assassination.

Constitutional Flaws Support the Aristocracy, Then Lead to Absolute Authority

The Roman constitution, though a great achievement, contained many serious flaws that contributed to the empire's demise. In addition to failing to give equal protection to all Roman citizens (thereby upholding the aristocratic class), it provided a pretext for senatorial corruption, set the stage for military coups, and allowed the appointment of dictators for extended periods of time.

Ten years after the deaths of the Gracchus brothers, the senate declined further. A war with North Africa found senatorial commanders being defeated on the field, then taking bribes to accept humiliating peace terms. Gaius Marius, a respected equestrian military leader, was elected consul and went to Africa, where he defeated the Numidians within two years and brought their king to trial and execution in Rome. Marius became a hero, and senatorial prestige took a severe blow. Marius then went on to save Italy from a barbarian invasion from two Germanic tribes, which made him even more popular (233). Marius had successfully combined military and political power. Marius reformed the Roman army into a professional one that served the interests of its commanders rather than the state, making way for the army to be used as a tool of politicians. To deal with the barbarian threats, Marius was elected consul six times, which violated the constitution.

War within the Senate

Internal war in the senate ensued. Cornelius Sulla, the quaestor of Marius, one of the best generals of the social war, became Marius's rival for leadership. Sulla's own followers in the senate had since outmaneuvered Marius and seized

control again; this was to Sulla's advantage, but there was a larger problem looming: senators were falling into debt because of their own excesses, and according to law, expected to face expulsion. In 88 BCE an activist tribune tried to expel those senators and bring the new allied citizens into power and substitute Marius for Sulla. Meanwhile, a revolt in Asia Minor had spread to Greece, with Athens and other towns threatening to break away from Rome. Dissatisfied subjects in the provinces joined the Asian revolt, and 80,000 Romans were massacred in a single day (in 237). Sulla had to put the domestic matter on hold until after he stopped the revolts in the provinces. After four years, Sulla put down the resistance and hurried back to Rome, but while he was away, his then-rival Marius had captured Rome and massacred Sulla's senatorial supporters.

Rise of a Dictator

Arguing that Rome had no legitimate government, Sulla made a plan to recapture Rome by declaring himself dictator, an emergency position appointed with the power to enact laws and take the necessary steps to end a crisis. When the danger passed, however, the dictator was expected to resign his office and restore normal government after six months. Sulla, however, was dictator for three years. He removed the equestrians from power, confiscated their lands, and killed many. He also increased the size of the senate and restored senators to the jury courts.

Rise of the Emperors

Sulla's opponents formed a government in exile in Spain and fought Rome on several fronts. After Sulla's death in 78 BCE, the senatorial, equestrian, and proletarian factions continued to fight. Ambitious generals from fights on all fronts sought power for themselves, including Julius Caesar, a supporter of Marius. In 60 BCE, Caesar, a former praetor of Spain, formed a triumvirate with the two other powerful generals at the time, Pompey and Crassus, to control Roman politics. Caesar became consul within a year and then added all of France, invaded Germany and Britain, and made a fortune in plunder with a five-year command. Caesar's victories in these regions provided him with a loyal army of soldiers more devoted to him than to the republic.

The rivalry between Caesar and Pompey led to Caesar becoming dictator. Pompey allied himself with the senate against Caesar. The senate moved to block Caesar's second five-year command in Gaul, as well as a consulship. Caesar refused to give up his command and a civil war broke out. Caesar captured Rome and ran out Pompey to Egypt, where he was assassinated. Within less than five years (49–44 BCE), Caesar extended his term as dictator to 10 years, then to life, and also served as the highest-ranking religious official as well. Although he instituted a beneficial tax reform and sponsored a successful colonization

program for the proletariat, Caesar irrevocably harmed the empire by claiming absolute power.

In what appears to have become an established practice by that time, Caesar was assassinated and a new civil war ensued. A second, but equally flawed, triumvirate was formed, and a battle ensued between two of the three, Antony, a general, and Octavian, Julius Caesar's grand-nephew and appointed successor. The empire was split for a time between the two men, but that process failed. Octavian eventually defeated Antony in the Battle of Actium in 31 BCE, and the civil war was ended. Octavian returned to Rome in 30 BCE and replaced the republic with a new form of government, the principate. Within the principate, the senate gained importance as a lawmaking entity with the power to elect officials. When Octavian named himself Emperor Caesar Augustus, a damaging turning point in the Roman Empire took place from which there was no turning back. From Augustus's time onward, many incompetent and ineffectual emperors would emerge to hasten the fall of the empire.

The danger of an emperor with absolute power was evidenced in the reigns of Caligula (41 BCE) and Nero (54 BCE). Caligula declared himself divine in 40 CE (Wise Bauer 2007: 725). The madness of Nero served as a dire turning point for Rome in particular. In 64 CE, a fire started in Rome, which quickly spread. Nero is rumored to have started the fire to make way for a new palace or just for his own amusement (Wise Bauer 2007: 729). Beyond the devastating destruction of Rome, Nero made another critical move that weakened the empire: he gave away Armenia. After allowing 3,000 victorious parthians, enemies of Rome, to march through the city to witness him hand over the Armenian crown, Nero lost the confidence of the Roman people. This disgrace, in addition to a series of gruesome personal atrocities carried out by Nero, emboldened the Praetorian Guard to carry out Nero's assassination.

The emperors were often mired in ceremony and grandeur, which served to isolate them from their subjects; this effect was "another of the fatal disunities that brought the Empire down" (Grant 1990: 100). Another intrinsic failure of the emperor model of government was the idea of "favor." The idea of a good emperor was one of a paternal patron. Subjects were expected to reciprocate in the form of deference, respect, and loyalty (Garnsey and Saller 1987). The emperor distributed benefits to favored groups, namely the aristocrats and the army. In the late first and second centuries, a growing number of officials served as mediators to the emperor, which was advantageous to the local elites. Collusion between imperial officials and local landholders allowed them to exploit the tenant farmers.

The late emperors were constantly surrounded by an extensive court that cut them off from the outside world. Proximity to the emperor meant great influence for members of the Imperial Court, which included a cabinet or council at its center; the aristocracy felt hostility toward the ruler's personal assistants. Wherever they were stationed, emperors were cut off from the world by scheming

courts. When the empire was split, the western emperors did not go out of their palaces, nor commanded the Roman army in wars. Both Honorius and Valentinian III preferred to remain in Ravenna, the new capital. This disconnect from the rest of the empire left the impression of a distant, absent government.

The army's entanglement in politics determined the outcome of a considerable number of emperors. Even after their arguably justifiable deposing of Nero by the Praetorian Guard, the guard capriciously switched allegiance seven months later to another potential emperor. At the same time, the army stationed at the Rhine River wanted yet another emperor, Vitellius, the commander of the forces in Germany (Wise Bauer 2007: 732). In the fourth century CE the army placed Valentinian I in power, and immediately afterward the soldiers demanded that he share power with a co-emperor. Fearing the possible death of a single emperor, the troops wanted to avoid chaos in that event. Valentinian took the western provinces and granted to his brother Valens the eastern provinces with equal legislative powers. All of the court services were duplicated in both capitals, a further drain on the economy.

The sense of a "unified" Rome with equal parts was shattered upon the death of Valentinian in 375. Valentinian's son Gratian failed to come to the aid of Valens, who was killed in battle against the Visigoths. Gratian made the critical decision of conceding Valen's successor Theodosius I most of the former western provinces in the Balkan peninsula. There had been an ongoing rift between the Latin West and the Greek East; Emperor Caesar Augustus believed that the Romans should maintain political supremacy over the Greeks (Grant 1990: 115). When Constantine the Great founded Constantinople as his capital in modern-day Turkey, it ushered in a new era where the East rediscovered its Greek heritage. From then on, the relations between the two empires unraveled and should be considered a major factor in the destruction of the weaker half, the West.

It can also be argued that due to a weakness in the imperial system, there was a break with tradition in the fifth century CE with the placement of a Hun into the highest command in the army, which led to civil war. By this time, Rome was dependent on barbarian soldiers for its defense, and this military had considerable influence on policy. Stilicho, appointed by Theodosius, was proof that "the Empire had to be defended not only against Germans without but against a German nation which had penetrated inside" (Bury 1992: 25). Stilicho's placement as the western empire's Master of Solders, or commander-in-chief, meant civil war, leading to the rebellion from the east.

The arrival of Hunnic tribes on the fringe of Europe in the fourth century who asked for asylum was a precipitous moment. Gratian unwittingly recruited these Huns into the army in 378 (Heather 1995). As a terrible result of the rift between the two governments, the Visigoth Alaric was able to enter Greece. Next, the eastern emperor Rufinus was assassinated, and suspicion fell on Stilicho. Stilicho was later assassinated at the request of senators to then western emperor Honorius.

The civil war between the western and eastern empires overshadowed the steady Hunnic penetration into the west. Civil war within the Hunnic empire left a final crisis for the western empire. Other barbarian groups broke free from the Huns after Attila the Hun's death in 453, and they started to press demands upon the western empire. Up to that point, Hunnic military power was being used by the western empire to contain immigrants on the borders. Cash-strapped, the western empire was unable to meet the barbarian army's demands. The eastern empire was both unable and unwilling to stop the ensuing invasions of the west; its provinces were doomed to become German kingdoms.

Imperialism Strains the Empire

The Military

In spite of the small proportion of power-holding aristocrats in comparison to the masses of poor commoners, the ruling class was able to maintain power practically in perpetuity; this hegemony was made possible by the military. Formal laws solidify the exercise of authority, and in the case of the early Romans, this power lay with the military-backed aristocrats. By the third century CE, internal instability led to 20 emperors recognized by the senate and 20 others laying claim to the throne with military support. The military became a destructive force in service to politicians.

Bureaucracy in Rome first appeared as a routine of the military discipline for which the Roman army was famous (Gibbon 1996: 10–17). The maintenance of the ancient world's most powerful army was the responsibility of a bureaucratic system able to recruit and financially care for it. The Roman military was hierarchical in structure, each with a fixed jurisdiction and a chain of command supported by tons of documentation. A Roman soldier was a professional who was specially trained, disciplined, and career minded. By the late republic, barbarians were allowed to serve and earn high rank, adding greatly to the number of soldiers to provide for.

As the bureaucracy advanced, emphasis was placed on merit and advancement. Dutiful soldiers could earn tenure, receive a bonus at retirement, and receive veterans' benefits. Along with the opportunities afforded Roman soldiers came the need for thousands of staffers, including accountants to handle payroll and messengers to check the ranks in the field.

An increasingly barbarian army irrevocably changed the social order of the Roman Empire. The civil strife weakened the empire and forced the emperors to resort to terror and compulsory laws. A new aristocracy emerged,

> which sprang up from the rank and file of the army . . . gradually produced a slave state with a small ruling minority headed by an autocratic monarch, who was commander of an army of mercenaries . . . the new army of the

second part of the third century was no longer the army of Roman citizens recruited from Italy and the Romanized provinces. . . . No sooner had this army recognized its own power at the end of the Antonine age, than it was corrupted by the emperors with gifts and flattery, and familiarized with bribery. (Rostovtzeff 1992: 31)

Each successive body of troops served completely at the pleasure of the emperor in power at the time. Eventually, the privileged classes found themselves at the mercy of a half-barbarian army. Later, in remote parts of the empire, the army served as the representative of Roman civilization. If a province was attacked, a soldier led the local defense and mobilized arms, men, and money (MacMullen 1992: 90). Civil authority was sanctioned by the government. Army garrisons often turned into cities themselves, but that extension would then require the building of official roads and bridges and require more troops, which was costly.

During the imperial era, the Visigoths became Rome's most successful enemy. The superiority of the Roman army was the result of the harnessing of military might for political purpose (Luttwak 1979). Having mastered large-scale warfare during the Punic Wars, Rome could not turn back to simpler times. Rome, embroiled in international warfare from then on, "never found the strength of will to lay down the sword. Her endless wars of conquest depleted her coffers, decimated her population, made enemies far and wide" (Bonta 2005).

Augustus's Pax Romana ended the war "industry." After he came to power, the army was supported by the considerable favor of the emperor. In the years of peace that followed, however, the army became a drain. The army rarely fought then, only in small barbarian skirmishes, leaving 250,000 to 400,000 idle mouths to feed. Augustus's division of the empire into armed and unarmed provinces weakened the empire, which left the region prone to social disorder and invasion.

By the end of the western empire in the fifth century, Rome had basically stopped producing its own soldiers. Conquered peoples enlisted in the Roman army as mercenaries. This practice was initially beneficial to the army, but when the soldier received his pension, he returned to his own nation, taking the pension with him, which would then contribute to the enrichment of peoples who would later sack Rome. Moreover, residents in "unprotected" provinces learned to defend themselves and became less reliant on the empire.

Later Negative Economic Effects Contribute to the Collapse

As previously mentioned, the ruling class objected to other forms of industry and discouraged the development of new technologies, which led to stagnation. For the wealthy, this land-based economy had to be maintained at all costs; any other economic improvements and innovations were viewed as a threat.

Competing industries such as mining were quickly shut down (Walbank 1992: 43), which would add to Rome's failure in later years. The economic market was locked in a limited circle of the upper and middle classes and the provincial armies, and the wider market of the peasantry was never tapped. In addition, the nobility's contempt for the artisan class (made up of primarily slaves and foreigners) overlooked a potentially lucrative part of the economy.

The end of slavery during Octavian's reign led to a shift in capitalistic activity within the empire. Freed slaves became citizens earning pay on village farms, which would eventually lead to urbanization. The gradual transfer of industry from the cities to the villages of former slaves with farming and other skills led to the decentralization of production. There were Romans "who preferred impoverished freedom to being anxious taxpayers" (Moorhead 2001: 27).

On the issue of taxes, one of the main concessions the patricians made to the plebeians was a tax reform that slightly reduced the impact of inherited status, which led to the rise of a capitalist class (Antonio 1979). However, the tax-paying proletariat saw no economic gains. A few hundred years after the republic was formed, Rome had gained control of the eastern Mediterranean. After the Punic Wars (264–149 BCE) with Carthage, the sea power, relegated to North Africa, the empire was extended to the eastern Mediterranean basin; this massive border expansion strained the republic. Higher taxes were required for the maintenance of the empire, and tax penalties crushed the poor.

There was little room for social mobility within the empire. Formal laws emphasized the nobility's exercise of authority over the poor. There were regulations regarding who qualified as having inherited status, and laws such as the Law of Debt required slavery for the poor (and their children) for failure to pay taxes. In contrast, aristocrats' tax delinquencies were overlooked or forgiven altogether. Poor farmers either abandoned their land or went to work for larger landholders, which reduced agricultural production, having a destructive effect in the later years of the empire. By the fourth century CE, the empire's ability to maintain its army was solely dependent on taxation. With taxes going directly to the military, the state was unable to benefit financially. A silver shortage also crippled the economy.

As Rome was primarily a land power, the maintenance of commerce over land was difficult. The upkeep of roads and the housing of traveling officials was expensive. A drop in long-distance trade occurred because products could be produced locally instead. At its height, the Roman Empire had nearly destroyed the local skilled networks and was relying on high-quality mass-produced goods such as pottery, boats, and wagons ordered from specialists hundreds of miles away. With the decline of the army and the lack of tax money to maintain roads, production and delivery were disrupted. The empire was reduced to that of "pre-Roman times, with little movement of goods, poor housing, and only the most basic manufactured items" (Ward-Perkins 2005: 137).

The End of the Empire

Rome, already collapsing from several internal problems, was unprepared and unable to regroup once the barbaric waves began. Roman civilization had kept Europe, the Middle East, and Africa cultured and prosperous for centuries, so it was a great loss when the western empire faced its demise. Other issues, such as the rise of Christianity and the growing power of the Holy Roman Church, played significant roles in the fall as well, but the larger issues already discussed are at the core of Rome's decline.

The civilized world could have possibly survived if the western and eastern empires had cooperated with each other, but the split between their governments was irrevocable. The resulting depopulation in the cities in the west was in part due to the artificial "newness" of the western way of living, compared to the east's longstanding traditions and agricultural practices long before they became romanized. The differences between west and east reached its apex in 476 CE when the last emperor of the west was deposed and his power fell to barbarian kings. After conflicts in the eighth and ninth centuries, the Christian churches split in 1054, when the pope excommunicated the patriarch of Constantinople and declared the Eastern Orthodox Church as heretical. The east was stronger than the west because wealth was more evenly distributed among its citizens, and there were fewer aristocrats (Ferrill 1992: 166).

No other empire since Rome has so fascinated the modern mind, due to its status as a forebear of modern civilization. The empire had unwittingly closed

---

Edward Gibbon and the Roman Pagan Ethic

One of the earliest and the best-known writers on the fall of Rome is British historian Edward Gibbon. Gibbon published his mammoth *The Decline and Fall of the Roman Empire* between 1782 and 1788. In it, Gibbon argued that Rome had become weakened over the centuries, as evidenced by the hiring of mercenaries to defend the empire, and thus had become too enamored of pleasure and the good life. Controversially, Gibbon argued that civic virtue fell out of favor with the rise of Christianity in the empire. Christianity's argument that life after death was more important than life on this Earth and its emphasis on pacifism weakened the traditional Roman emphasis on the military and the duty of its citizens to support the preeminence of Rome. Because of his assertions, and the prominence of church historians at the time, Gibbon was called a "paganist" and was accused of disputing the doctrine of the Roman Catholic Church, whose history is intimately tied with that of the Roman Empire after the Emperor Constantine. Some have argued that Gibbon's writings were the beginnings of the professionalization of historians, as he looked to the contemporary sources to shape his interpretations rather than to accepted church interpretations to shape his recounting of the events and their meanings.

in upon itself, subjugated its allies, and then presented to its democracy the contradictory idea of an emperor's authority. The legacy of the rule of Roman dictators and emperors will live in perpetuity, leaving the world forever changed. The governmental and architectural achievements of Rome teach us much, but the failures of Roman society teach even more.

References and Further Reading

Antonio, Robert J. "The Contradiction of Domination and Production in Bureaucracy: The Contribution of Organizational Efficiency to the Decline of the Roman Empire." *American Sociological Review* 44, 6 (1979): 895–912.

Bonta, Steve. "Lessons of Rome: the Rise and Fall of the Roma Republic Provides Lessons that Hint at Flaws in Modern Political Policies." *New American* (2005). Available at http://findarticles.com/p/articles/mi_m0JZS/is_4_21/ai _n25105131/ (accessed August 3, 2010).

Bury, John B. "Decline and Calamities of the Empire." In *The End of the Roman Empire: Decline or Transformation?*, 3rd ed., edited by Donald Kagan. Boston: Heath, 1992.

Ferrill, Arther. "The Empire Fell." In *The End of the Roman Empire: Decline or Transformation?*, 3rd ed., edited by Donald Kagan. Boston: Heath, 1992.

Garnsey, Peter, and Richard Saller. *The Roman Empire: Economics, Society and Culture*. Berkeley: University of California Press, 1987.

Gibbon, Edward. *The Decline and Fall of the Roman Empire*. Edited by David Widger. Salt Lake City, UT: Project Gutenberg, 1996.

Grant, Michael. *The Fall of the Roman Empire*. New York: Collier Books, 1990.

Heather, Peter. "The Huns and the End of the Roman Empire in Western Europe." *English Historical Review* 110:435 (1995): 4–41.

Jones, Tom B. *From the Tigris to the Tiber: An Introduction to Ancient History*. Belmont, CA: Wadsworth, 1989.

Luttwak, Edward N. *The Grand Strategy of the Roman Empire: From the First Century A.D. to the Third*. Baltimore: Johns Hopkins University Press, 1979.

MacMullen, Ramsay. "Militarism in the Late Empire." In *The End of the Roman Empire: Decline or Transformation?*, 3rd ed., edited by Donald Kagan. Boston: Heath, 1992.

Moorhead, John. *The Roman Empire Divided, 400–700*. New York: Longman, 2001.

Rostovtzeff, Michael I. "The Empire during the Anarchy." In *The End of the Roman Empire: Decline or Transformation?*, 3rd ed., edited by Donald Kagan. Boston: Heath, 1992.

Walbank, F. W. "Trends in the Empire of the Second Century A.D." In *The End of the Roman Empire: Decline or Transformation?*, 3rd ed., edited by Donald Kagan. Boston: Heath, 1992.

Ward-Perkins, Bryan. *The Fall of Rome and the End of Civilization*. New York: Oxford University Press, 2005.

Wise Bauer, Susan. *The History of the Ancient World: From the Earliest Accounts to the Fall of Rome*. New York: Norton, 2007.

CON

Although many causes can contribute to the fall of an empire, which is in fact the result of a process, even recent historical evidence shows that empires lose momentum when their territorial expansion is over, and that gives way gradually to territorial contraction or internal collapse. The collapse of the Soviet Union was preceded by losing the war in Afghanistan, which was its first loss during its history of unabated growth in political influence and military might. In the case of the British and French empires, the world wars contributed to weakening of their control, and the rise of nationalism and anticolonial movements were spurred by the nationalism of the empire-building nations. The German Third Reich came to an end after losing World War II, and the circumstances weren't very different in the case of the Roman Empire.

Some argue that the Roman Empire lived on as the Frankish king Charles the Great (Charlemagne) declared himself to be Holy Roman emperor in 796, but the same could also be said about the German–Roman Empire of the Middle Ages, with their desire to legitimize power in the name of antiquity, historic rights, or even divine rights. But the empire ended when none claimed or was able to claim the title, and this is valid for any empire since (Chinese emperor, Russian czar, Ottoman sultan, German kaiser, Austro-Hungarian monarch), whether there was a military defeat or not. The formal end of the Roman Empire corresponds with the time in which the empire and the title "emperor" no longer had value, therefore rendering it senseless in practice. We can rightly assume that a decisive military defeat and the occupation of the capital, Ravenna, and the complete partition of the former territory are decisive in defining the end of an empire. This may lead to the abolition of the main social–political features of the empire for a time, but its political existence is over forever.

The Beginnings of the End

The empire reached the peak of its territorial expansion in 106 CE, by conquering Dacia, but due to the military pressure of the barbarian tribes and peoples coming from the east, it had to give up and evacuate it in 271 CE, only after 160 years of

conquest. The first signs of its impending collapse were shown during the so-called crisis of the third century (234–284), a period of political anarchy. The empire almost collapsed due to the greatest Gothic invasion seen to date, which hit the Balkans. The Goths' seaborne allies, the Heruli, supplied a fleet, carrying vast armies down the coast of the Black Sea, where they ravaged coastal territories of Thrace and Macedonia, while land forces crossed the Danube in Moesia (roughly modern-day Bulgaria). Similar Gothic invasion was hitting the province of Pannonia, which lead to disaster. The empire struck back in 268, with Emperor Gallienus wining some important victories at land and sea, but it was his successor Claudius II who finally defeated the invaders at the Battle of Naissus (modern-day Nis, in Serbia) in 268. This was one of the bloodiest battles of the third century; the Goths allegedly left 30,000 to 50,000 dead on the field. The reign of Emperor Diocletian (284–305) regained control and saved the empire by carrying out substantial political and economic reforms, many of which would remain in force in the following centuries.

Erennio Etrusco, son of Decius and leader of the Roman army in the battle against the Goths at Abrittum in 251 CE. (Vanni/Art Resource, NY)

But by fourth century the heyday of the Roman army was almost over. Two main forces started to shape the destiny of the western empire: the Huns and the Germanic tribes. The main Germanic tribe, that of the Visigoths, who had fled from the Huns, at first had been allowed to settle within the borders by Valens, to become peasants and soldiers of the empire. As is usual they were mistreated by the local Roman administrators and rebelled, leading to the first war with the Visigoths (376–382). The Goths crossed the Danube again in 378 and defeated the Roman army at Adrianopole and subsequently settled in the Balkans, becoming known as eastern or Ostrogoths. This battle was of great political and strategic importance and the heaviest defeat of the Romans for four centuries. The Roman defeat left a large and hostile foreign force within the frontiers of the empire. This was followed by the reorganization of the army under the reign of Theodosius I (379–395), the last emperor succeeding to reunite under his authority the western and eastern halves of the empire. By

doing so he outlawed paganism and made Nicaean Christianity the state religion, a move that also contributed to the collapse of the empire, lowering the fighting spirits of the troops.

Even before this event, the penetrations of barbarian forces was not uncommon, but after it, it became almost permanent. After the Goths, the Huns also crossed the Danube many times, and Attila the Hun conquered among others Naissus in 443, with battering rams and rolling towers, his military sophistication showing that his army had also adopted Roman methods of warfare.

The fall of the west wasn't without long lasting consequences, and that was the weakening of the east. The east also started to fall during the Gothic war launched by the eastern emperor Justinian in the sixth century, which aimed to reunite the empire again, but that eventually caused the most damage to Italy and strained the eastern empire completely in military terms. Following these wars, the Italian cities had fallen into severe decline, Rome itself being almost completely abandoned. A last blow came with the Persian invasion of the east in the seventh century, immediately followed by the Muslim conquests, especially of Egypt, which curtailed much of the key trade in the Mediterranean, on which Europe depended. Byzantium became a Greek empire, centered mostly nowadays in Greece and Turkey (Anatolia), and finally shrunk to the city of Constantinople itself, which fell ultimately in 1453.

Causes Leading to the Weakening of the State Supporting the Army

The first major problem that caused the weakening of the state supporting the army comprised the collapse of the Roman tax system; the looting economy ended with the territorial expansion and the weakening drive of military glory due to the lack of military successes. The first major reduction of real income appeared after the Sassanid wars in the third century, which stabilized the borders on the east, but also led to two negative long-term impact trends. First, the incentive for local officials to spend their time and money in the development of local infrastructure disappeared. Public buildings from the fourth century onward tended to be much more modest and funded from central budgets, as the regional taxes had dried up. Second, the landowning provincial literati shifted their attention away from provincial and local politics to the imperial bureaucracies, where tax money was concentrated. As archaeological evidence shows, the Germanic tribes on the empire's northern border increased their material wealth, due to their contact with the empire, and that in turn had led to disparities of wealth sufficient to create a ruling class and oligarchs capable of maintaining control over far larger groupings of their people than had previously been possible, therefore becoming significantly more formidable adversaries. The settled barbarians, however, due to high taxes, Roman prejudice, and government corruption, gradually turned against the empire and began looting and pillaging throughout the eastern Balkans.

Disease, Causing Population Decline

Archaeological evidences show that Europe continued to have a steady downward trend in population, starting as early as the 2nd century and continuing until the 7th century. This is largely due to the epidemics of smallpox and measles, which swept through the empire during this period, ultimately killing about half of the entire population. The severe drop in population left the state apparatus and army too large for the population to support; therefore, the depopulation of Italy favored Germanic expansion. The eastern half survived, however, due to its larger population, which was sufficient for sustaining an effective state apparatus. The European recovery may have started only once the population had gained some immunity to the new diseases. The ravages of disease didn't end with this. Later, the plague of Justinian may have been the first instance of bubonic plague, and it was so devastating that it helped in the effort of the Arab conquest of most of the eastern empire and the whole of the Sassanid Empire (most of modern Iran, Iraq, and Afghanistan).

The empire underwent major economic transformations. The economy of the empire was a kind of complex market economy in which trade was relatively free. Tariffs were low and laws controlling the prices of foodstuffs and other commodities had little impact because they did not fix the prices significantly below their market levels. On the other hand, the Romans had no budgetary system and thus wasted whatever resources they had available. The economy of the empire was basically a plunder economy, based on looting existing resources rather than producing anything new. Material innovation and technological advancement all ended long before the final dissolution of the empire, and as a consequence few exportable goods were produced. The economy, based upon slave labor, was inefficient and precluded having a middle class with purchasing power that could support local businesses. Roman territorial expansion was also a consequence of decreasing Roman agricultural output and increasing population. By conquering their neighbors, the Romans appropriated their energy surpluses (metals, grain, slaves, etc.). The booties stopped with the end of territorial expansion, but the large empire was costly to uphold, in order to maintain communications, garrisons, and civil government on a functional level. With the cessation of tribute from conquered territories, the full cost of the Roman military machine had to be borne by the citizenry, and this was increased even more by the pomp costs of the emperors. Eventually this cost grew so great that any new challenges such as invasions and crop failures could not be solved by the acquisition of more territory, and at that point, the empire fragmented into smaller units and became ripe for collapse.

By the third century beginning with the reign of Nero, the monetary economy of the empire had collapsed, because Roman mines had peaked and output was declining and territorial expansion stopped. Since mines of all commodities

were being depleted, this led to the debasement of the currency and subsequent inflation. The lack of metals caused a decline of Roman technological and economic sophistication, while the inflation required that emperors adopt price control laws that resulted in prices that were significantly below their free-market equilibrium levels, transforming the empire into a kind of state socialism. These artificially low prices led to the scarcity of foodstuffs, particularly in cities whose inhabitants depended on trade in order to obtain them.

To counteract the effect of inflation on state revenues, the monetary taxation was replaced with direct requisitioning from the farmers. Individuals, most of them slaves, were forced to work at their given place of employment and remain in the same occupation. Farmers became tied to the land, as were their children, and similar demands were made on all other workers and artisans as well. In the countryside, people attached themselves to the estates of the wealthy in order to gain some protection from state officials and tax collectors. These estates, the beginning of feudalism, operated as much as possible as closed systems, providing for all their own needs and not engaging in trade at all. Workers were organized into guilds and businesses into corporations called *collegia*, and both became de facto organs of the state, producing for the state. There was a decline in agriculture, and land was withdrawn from cultivation, in some cases on a very large scale, sometimes as a direct result of barbarian invasions. However, the chief cause of the agricultural decline was high taxation on the marginal land, driving it out of cultivation. High taxation was spurred by the huge military budget and was thus indirectly the result of the barbarian invasion.

Environmental causes also played a significant role in the weakening of the state. The complex Roman society depleted its resource base beyond levels that are ultimately sustainable. The existence of large nonproductive masses, such as the army and bureaucracy, its establishment and running of infrastructures, which require substantial "energy" subsidies, overburdened the tax base, which collapsed as gradual environmental degradation caused crop failures and subsequent population decline. Deforestation and excessive grazing led to erosion of meadows and cropland. Deforestation had significant military causes as well, ensuring that the forests could not provide cover and camouflage for the enemies, usually attacking in smaller and less organized groups but using the advantages of the terrain. Shipbuilding was also a major contributor to deforestation. At times of war, hundreds of ships could be built within a month, leading to scarcity of timber in the immediate areas of shipbuilding centers.

The intensive agriculture needed to maintain a growing population required irrigation, which in turn caused salinization, resulting in fertile land becoming nonproductive and eventually increased desertification in some regions, mainly in Africa, Hispania, the Balkans, Anatolia, and in the Middle East. Many animal species became extinct, as the empire practiced large-scale and systematic environmental destruction. At shows held at the Coliseum in Rome there may have

been up to 5,000 lions, bears, and wolves killed annually, leading to their extinction in all of western Europe and even in northern Africa and the Middle East (the Atlas lion being the favorite beast of the gladiator shows).

Lowland areas and areas close to water transports were highly urbanized first, but as population increased along merchant routes, the environment underwent drastic degradation as pollution from the burning of fuel wood filled the air and smelters that used wood as fuel transmitted heavy metals into the atmosphere. Urbanization and the resulting degradation of the environment ultimately weakened the Roman Empire, reducing its capacity to supply the necessary raw material and money for defense, which in turn resulted in the considerable weakening of the army.

The Weakening Army

The Roman military can be viewed as one of the greatest armies ever. However, from around the end of the fourth century, the military began to deteriorate continuously and irreversibly. Its causes are manifold. Perhaps the most important was the abundance of internal conflicts in which the military fought, split between rival contenders to the throne of the empire. The endless civil wars between factions of the Roman army fighting for control eroded the political capacity to maintain the army at superior organizational, tactical, and armament levels. On the contrary, some argue that the army still remained a superior force compared to its civilized and barbarian opponents, as is shown in its victories over Germanic tribes at the Battle of Strasbourg in 357 and in its ability to hold the line against the Sassanid Persians throughout the fourth century.

In spite of these instances, the prevalence of civil wars during the fourth and fifth centuries, when the Roman army was forced to fight, caused a constant drain of resources that might have been used against external enemies.

Treason and the killings of army generals, always a decisive factor in military defeat, became omnipresent. A critical event was the murder of Majorian (457–461) in 461, denying him the chance of recapturing Africa and holding the western empire together. After Majorian's murder, western imperial unity finally dissolved. Aegidius in Gaul and Marcellinus in Dalmatia refused to accept the new Emperor Libius Severus (461–465), who rose to power thanks to Ricimer's murder of Majorian. Another reason the army was weakened was the mixing of the ranks of the legions with barbarian mercenaries, with the army becoming basically a mercenary army recruited mainly from neighboring Germans. These barbarians were willing to join the army, either to stop the Romans from invading their territory or because they wanted Roman citizenship if they served for a certain number of years, as is also the case in modern mercenary armies, such as the French foreign legion.

The process of admittance was started by the emperors Diocletian (284–305) and Constantine I (305–337), who split the army into border and mobile

components. The border troops became soldier-farmers and declined rapidly in efficiency, though they were still paid. The weakness of the border troops meant that emperors needed more mobile troops, so they expanded the army. This in turn increased the number of recruits needed, while the simultaneous reluctance of landowners to lose scarce workers led to the recruitment of the militarily inferior barbarians. The funds for the army were low; therefore, it had problems in recruiting Roman citizens, resorting rather to the cheaper but less trained and conscious barbarians, which had the result of decreasing its fighting efficiency. The military hired barbarians because Romans were unwilling to join, some even willingly cutting their fingers to avoid being drafted. As a result, the senate made service compulsory, but even this did not help. As a result, there were more Germans in the army then Romans, and this is why the barbarians were able to control the army. This Germanization and the resultant cultural dilution or barbarization led to lethargy, complacency, and loyalty of the legions to the Roman commanders, instead of the Roman government, and a surge in decadence among Roman citizenry.

The government promised higher cash rewards to those who joined the army, necessary to render the service attractive, but as the army became dominated by its commanders, which formed militias, one of them of Orestes, they became crucial in destabilizing the empire. The setup of a mercenary army was a direct consequence of the decline in military spirit and of depopulation in the old civilized Mediterranean countries, and the moral decay was due to the affluence experienced before, including a decline in patriotism and loyalty among soldiers. The Germans in high command had been useful, but they were also dangerous as the case of Odoacer has shown clearly. Since the barbarians could not be trusted, any successful army tried to avoid too much interaction with the enemy since its influence would become subject to collapse. The barbarians realized the importance of the army, started to take advantage of its unorganized nature, began to infiltrate the empire, and finally took over the empire.

The Romans could maintain their power only by continuously making and changing alliances with their enemies, pursuing a *divide et impera* (Latin for divide and rule) politics among the barbarians. In fact the Roman army wasn't able to destroy the barbarians that were attacking, and they couldn't keep other tribes out. These frequent alliances with the barbarians convinced the latter of the vulnerability of the empire and of any of its emperors. As Alaric was used against the western Romans, Maximus, the main plotter against Aetius, disillusioned by the emperor and his former ally Heraclius, used the Huns to get revenge on them, so two Hun friends of Aëtius, Optila and Thraustila, assassinated both Valentinian III and Heraclius in 455. Since most of the Roman soldiers present at the scene of stabbing were faithful followers of Aëtius, none made an attempt to save the Roman emperor. Prior to that, for example, by the time of emperor Caligula's assassination in 41 CE, his loyal Germanic guard

responded with a rampaging attack on the assassins, conspirators, innocent senators, and bystanders alike.

Constantine I built the new eastern capital of Constantinople and transferred the capital there from Rome in 330, the east being promoted to the core of the empire. This occurred because Greek-speaking Christians—after years of persecution—had taken over the Roman Empire. Thus, what little available resources of metal they had were used to save the new capital city of the Roman Empire and its adjacent provinces of Greek-speaking Christian Anatolia. As a result, the Greek-Christian Romans drove all the Germanic invaders toward the Latin West, which had been demoted to the periphery. If the capital of the Roman Empire had not been transferred, then the authorities would have driven the Germanic invasions toward Anatolia, and the west could have been saved. Although the institution of the empire itself was not at fault, poor leadership played a significant role in the military failure of the empire.

The impact of Christianity was significant in other ways. Not upheld by conscious warriors, no longer proud of their cause and virtues, the army's morale had inevitably fallen, leading to a decrease in fighting virtue. The many losses the Romans had suffered further decreased the morale among soldiers. Christianity transformed its status from an oppressed and brutally persecuted religion into the official dominant religion of the empire, during the reign of Constantine. Consequently, most of the people became Christians, including all the Roman emperors after Constantine, except Julian. As the Christian philosophy favored pacifism, Christians were unwilling to fight, expecting rewards in heaven rather than from their daily lives. The more Christian the soldiers became, the more they lost their martial virtue, which was considered a sin in itself, and resulted in the lack of moral superiority.

The lack of technical superiority of the Roman army was another evident cause of the empire's collapse. The Roman army, based on infantry, lacked mobility and the adequate cavalry capable of securing that. Although it had cavalry, it could not match the completely mobile armies of their enemies on horseback. The horseshoe was invented by the Barbars, and their mobile army altered warfare from infantry to cavalry dominance for the next thousand years, a trend marked by the second Adrianople defeat in 378 CE. The lack of funds led to an ever-poorer quality of the weaponry and armor supplied to the troops, which in the end became so obsolete that enemies had better armor and weapons as well as larger forces. The decrepit social order offered so little to its subjects that many saw the barbarian invasion as liberation from their onerous obligations to the ruling class.

The Unfavorable International Situation

The empire faced ever growing external threats. The first was the emergence of the Sassanid Persian Empire in Iran (226–651), which resulted in the withdrawal

of the Roman legions from the line of the Euphrates and also from much of Armenia and Kurdistan. Indeed, 20–25 percent of the overall military might of the Roman army and up to 40 percent of the troops under the eastern emperors confronted the Sassanids for about half a century. The constant pressure exerted by invading people caused a domino effect—the pressure on people who are living far from the empire resulted in sufficient pressure on people who lived on the empire's borders to make them contemplate the risk of full-scale immigration into the empire. Thus the Gothic invasion of 376 was directly attributable to Hunnish advancements around the Black Sea in the previous decade. In the same way the Germanic invasions across the Rhine in 406 were the direct consequence of further Hunnish incursions into Germania, so the Huns became deeply significant in the fall of the western empire long before they themselves became a military threat. The 395 division of the empire, between Arcadius and Honorius, the two sons of Theodosius, was also very unfortunate, since it ended the strategic depth of the empire, allowing fewer transfers of money and troops between the two parts. With longer borders to defend and fewer resources, some historians argue that the western collapse was inevitable after 395.

Evidence that Supporting the Military Causes Lead to the Fall of the Empire

The previous military defeats were a prelude to the final collapse of the empire. Its major landmarks are the battle of Adrianople with the Goths in 378; the crossing of the Rhine by the Germanic tribes and Alans in 406; the first sacking of Rome after 800 years in 410 by the Goths; the loss of Africa in 426 to the Vandals; the battle of Catalaunum in 451 with the Huns and the subsequent Hunnish incursion into Italy; the second sacking of Rome by the Vandals in 455; and the final push by Gothic leader Odoacer in 476 CE.

At the second Battle of Adrianople in 378, some 10 km from Edirne in modern Turkey, though the Roman army had a clear numerical superiority (20,000–40,000 opposed to 12,000–15,000 of Goths, according to estimates) and also had Emperor Valens on the battlefield, their effort was disorganized and undisciplined. This Roman army consisted of heavy infantry, various archers, and cavalry, but it attacked the circular Gothic coach camp (wagon lager) too late and in disarray. In the meantime, the Gothic heavy cavalry encircled the Roman infantry, and with its Gothic and Alan infantry pushing from the front, it caused the whole Roman army to fall into a disorganized rout, leading to the massacre of two-thirds of it, including the emperor himself. The causes of the defeat were manifold, including redeployment problems and simultaneous attacks on other fronts, leading to insufficient number of troops in the three Roman armies, which theoretically fought together.

A peace agreement was forged with the Goths in 382, in which the new eastern emperor, Theodosius I, had recognized their claim to the province of

Thrace. The eastern emperors, as would be the rule in the Byzantine Empire, used deception and lies to weaken their enemies, since they no longer had adequate military fighting power. As an example of this tactic the invader Visigoths became allies and used this against the western Roman army under Eugenius and Arbogast, which were supporting the western contender that had been rejected by Theodosius. Alaric the Visigoth accompanied Theodosius's army in invading the west in 394. At the Battle of the Frigidus, Theodosius had explicitly ordered the Goths to charge the usurpers army before he engaged his soldiers, with the intent of weakening both the Visigoths and the western Romans. This led to the sacrifice of the Visigoths with approximately half of them dying, which enabled Theodosius to win the battle. This convinced Alaric that the Romans sought to weaken the Goths by making them bear the brunt of warfare in order to completely subjugate them after Theodosius died in 395. Alaric soon resumed hostilities against the eastern empire.

Fearing the Visigoths, Emperor Diocletian moved the capital of the western empire from Mediolanum (modern Milano) to Ravenna in 286, which was better located strategically. In the meantime, Alaric made several attempts at invading Italy, but was halted by Flavius Stilicho and decisively defeated at the Battle of Pollentia and later in the Battle of Verona, but the west was unable to succeed alone, so it finally proposed an alliance with the Visigoths to help reclaim Illyricum. However, when the Vandals and Sueves crossed the Rhine and invaded Gaul, the invasion was called off and Alaric was left with the expense of preparations for the campaign. Stilicho persuaded the Roman senate to reimburse Alaric, but the fiasco had sown resentment in both the Romans and in Alaric's Goths. These events were followed by more violence on the part of the Roman army, this time aimed at the barbarian soldiers and slaves in Italy, many of them who had been captured by Stilicho in his earlier wars. Around 30,000 escaped Italy and fled to strengthen Alaric's army. This was paralleled at the end of the fourth century when small barbarian tribes who had opposed the early empire now banded together to form more powerful confederations such as the Goths, Franks, and Alamanni. The Danube had been crossed many times before 406, but the Rhine had never been, at least decisively. When the Huns arrived in 360 and drove the Goths into the empire, the Roman army near Constantinople, the capital, was defeated. From then on, the Romans could not destroy these Goths, and after one group of barbarians had entered the empire, the Romans could not muster the military strength to keep others out. So on December 31, 406 (or 405, according to some historians), a mixed army of Vandals, Suebi, and Alans crossed the frozen river at Moguntiacum (modern Mainz) and began to ravage Gaul, some moving on to Hispania and Africa, with the empire losing de facto control over most of these lands forever. These invaders went on to settle all over the western empire: the Visigoths, Alans, and Suevi took land in Spain; the Vandals in Africa; and the Burgundians, the Visigoths and Franks in Gaul. Some

historians are doubtful about the increased power of these nations and claim that emperors like Diocletian, Constantine, Constantius II (337–361), and Valentinian I (364–375) kept the barbarians beyond the borders. But the process continued after the fall of the empire, such as the Saxons invading Britain and at the end of the fifth century, while the Ostrogoths occupied Italy.

The sieges and sackings of Rome proved to be the decisive military weakness of the empire. The decisive Visigoth political interfering at the highest level became visible soon after they invaded Italy and laid the first siege to the city in late 408. Rome was being affected by starvation and disease and became desperate, so the senate sent several envoys, including the pope, to Ravenna to encourage Emperor Honorius to strike a deal with the Goths. When this didn't happen, the Roman senate offered Alaric a nice amount of gold, silver, silk, and pepper in exchange for lifting the siege. This showed that the Romans lacked a sufficient army to defend even the heartland of the empire. This military weakness proved decisive for Alaric, who wanted the provinces of Rhaetia and Noricum (roughly present-day Austria, Switzerland, Slovenia, and southern Bavaria) as a home for the Visigoths, and the title of Magister Militium, but the emperor refused this and tried to sneak a force of Illyrian soldiers into Rome. The army was intercepted by Alaric and, outraged by the insult, besieged Rome for a second time, destroying the granaries at Portus. Faced with the return of starvation, the senate surrendered again, this time under pressure from Alaric, who appointed Priscus Attalus as a rival emperor. So the barbarians actually decided who would be the emperor. Alaric was made Magister Utriusque Militium and his brother-in-law Ataulf, the commander of the cavalry, Comes Domesticorum Equitum. They marched toward Ravenna to depose Honorius, who was ready to surrender when an army from the eastern empire arrived to defend Ravenna. As another blow, Heraclian, the governor of the African coasts, cut off Rome's grain supply, threatening the city with another famine. Alaric wanted to send Gothic troops to invade Africa and secure food for Rome, but Attalus refused, probably fearing that the Goths would seize Africa for themselves. In response, Alaric ceremonially deposed Attalus and reopened negotiations with Honorius.

When Alaric was on the verge of an agreement with Honorius, his forces were attacked by Sarus, a fellow Gothic commander who was allied to Honorius and who had a blood feud with Ataulf. Alaric returned to Rome and laid siege to it a third time. On August 24, 410, the slaves opened Rome's Salarian Gate, and the Visigoths poured in and looted for three days. Because they had already been converted to Christianity, it was not a particularly violent looting, but it still had a profound effect on the city, since it was the first sack after almost 800 years of successful defense, after the victory of the Gauls lead by Brennus in 387 BCE. This sack proved to be a major historical landmark, as many great edifices were ransacked, including the mausoleums of earlier Roman emperors. As a result, tens of thousands of citizens fled, many into Africa. After raiding Greece and

invading Italy, the Visigoths settled elsewhere in the empire, founding the Visigoth kingdom in southern Gaul and Hispania.

Although Roman and most Western sources claim that the Roman–Visigoth army defeated the Hunnish alliance on the fields of Catalaunum in 451, the battle was rather indecisive. The Romans built on an alliance with the Visigoths, whom they had defeated earlier under the leadership of Aetius in the Battle of Narbonne in 436. This conclusion of indecisive outcome in 451 is supported by the facts that the Huns left the fields the second day completely organized, the king of the Visigoths was killed during the battle, and the Romans did not pursue them. Aetius himself had grown up among the Huns as a royal hostage and knew better than anyone else that the victory, if we may call like that, wasn't decisive. This was also reinforced by the fact that shortly after the battle, in 452, Attila again entered Italy itself looking to marry the daughter of the emperor. Attila's friendship with Aetius and his respect of his enemies and for the Roman culture caused him to behave extraordinarily civilized, even more than other Roman commanders of the time, and he withdrew from the gates of Rome, probably under some Christian influence but specifically on the request of pope who was absolutely mesmerized by this behavior and conferred him the title of "Scourge of God." It is evident that the Huns could have ransacked Rome in 452 as they did in other Roman towns because Roman defense was nowhere to be found. But what was delayed for the time would eventually happen.

The Vandals invaded the African provinces of the western empire (Mauretania) in 429, mostly completing it by 439. This had severe financial and strategic consequences; it not only ended the western control of these wealthiest provinces, but it also exposed the Mediterranean to pirate raids. This was shown clearly during the second sack of Rome, which was performed by the Vandals in 455, when a Vandal fleet led by King Geyserik could sail up the Tiberis from Carthage, showing that there was no Roman navy present or capable of fighting, nor any field army to resist it. In opposition to the forgiving Huns or moderate Visigoths, the Vandals had been looting the city for 14 days, taking also shiploads of slaves, giving birth to the term "vandalism," meaning useless and senseless destruction. The Vandals then moved to Africa, but the western emperor Majorian was unable to retake it in 461. During his preparations, the news got out to the Vandals, who took the Roman fleet by surprise and destroyed it. A second naval expedition against the Vandals, sent by Emperors Leo I and Anthemius, was defeated in 468.

In 476, Odoaker, a leader of the Germanic *foederati* was promised land in Italy by militia leader Orestes for fighting against Nepos, the emperor. The *foederati* were Germanic troops under their chieftain's commands. Julius Nepos was nominated emperor by the eastern emperor Zeno and considered to be the legitimate emperor by Zeno's rival Basiliscus. As Nepos fled to Dalmatia, the promissor Orestes placed his son Romulus on the throne, naming the emperor Augustus, but failed to keep the promise he had made to Odoacer. Outraged by

A Philosopher-Solider Views the End of His Son and the Empire

Written in 413 CE, this letter by Synesius of Cyrene, a philosopher-soldier, demonstrates his view that Christianity did not weaken the Roman military ethic, but, in the life of his son Marcellinus, enhanced it.

To the General

Praise is the reward of virtue, which we offer to the most illustrious Marcellinus at this moment when he is leaving his post, at this moment when suspicion of every flattery is in abeyance. When he arrived here, he found our cities attacked from without by the multitude and rage of the barbarians, from within by the lack of discipline of the troops and the rapacity of their commanders. Marcellinus appeared in our midst as a god. He vanquished the enemy in a single day's fighting, and by his continual alertness he has brought our subjects into line. He has thus out of both calamities brought peace to our cities.

Nor did he claim any of those profits that usage has made to appear lawful; he has not plotted to despoil the rich or ill-treat the poor. He has shown himself pious towards God, just towards his fellow citizens, considerate to suppliants.

On this account a philosopher priest is not ashamed to praise him, a priest from whom no one ever received a testimonial bought by favor. We wish that the courts of law also were present with us, so that, collectively and individually, all we inhabitants of Ptolemais might have presented him in return with such a testimonial as is in our power, however inadequate, for words are somehow far inferior to deeds. I would most willingly have made a speech on the occasion in behalf of us all.

But since today he is beyond the frontier, we wish at all events to dedicate to him our testimony in the form of a letter, not as those from whom a favor is solicited, but as those who have solicited one.

Source: "Letter 62: To the General: a Farewell." *Letters of Synesius of Cyrene*, translated by A. Fitzgerald. London: 1926.

this, Odoacer defeated him in the Battle of Ravenna and entered the capital and forced the child emperor, Romulus Augustus, to abdicate and killed Orestes. Although Odoacer recognized Nepos until its death in 480, he put an end to the western Roman empire by sending the imperial insignia to Constantinople. The significance of this is reinforced by the fact that he didn't proclaim himself as an emperor, although he could have. Strictly speaking he wasn't allowed to hold the title of emperor because he wasn't a Roman citizen, therefore he asked Zeno to become formal emperor of the entire empire, and in so doing he legalized Odoacer's own position as imperial viceroy of Italy. Odoacer could also have chosen himself a puppet, since he legally kept the lands as a commander of the eastern empire, in the name of Zeno. Zeno recognized him as imperial viceroy, deposing at the same time Nepos, the last legitimate western emperor,

who was murdered by his own soldiers in 480, putting an official end to the already defunct western empire.

By keeping the Roman administration, senate, and most of the taxes in place for a time, Odoacer was accepted for good not only by his conational Germans, but also by the neo-Latin speakers, becoming the first German-Italian king of Italy. The Germanic *foederati*, the Scirians, the Heruli, as well as a large segment of the Italian Roman army, proclaimed Odoacer "king of Italy" (*rex Italiae*). Later, Zeno, concerned with the success and popularity of Odoacer, started a campaign of slander against him, inciting the Ostrogoths to conquer and take back Italy from him, which happened with Zeno's approval, but instead of returning the western empire, they founded their independent Ostrogothic kingdom in Italy in 493, under the rule of their king Theodoric the Great.

Although there were other troublesome changes that took place in the Roman Empire during the fourth and fifth centuries, the true cause of the empire's downfall was primarily military in nature. The empire had reached its military and territorial zenith, but had become so overburdened by its own needs in terms of soldiers (necessitating the heavy use of mercenaries), material (greatly impacting the Roman civilian population), and political influence (becoming part of the imperial intrigue rather than separate from it) that it essentially collapsed upon itself, allowing the barbarian armies an opportunity of which they took full advantage. With the influx of Germanic tribes in the western empire, the age of Roman power truly had come to an end.

References and Further Reading

Cameron, Averil. *The Mediterranean World in Late Antiquity, CE 395–600*. London: Routledge, 1993.

Davis, Paul K. *100 Decisive Battles from Ancient Times to the Present. The World's Major Battles that Shaped History*. New York: Oxford University Press, 1999.

Elton, Hugh W. *Warfare in Roman Europe: CE 350–425*. New York: Oxford University Press, 1996.

Ferrill, Arthur. *The Fall of the Roman Empire: The Military Explanation*. New York: Thames and Hudson, 1986.

Galsworthy, Adrian. *The Complete Roman Army*. London: Thames and Hudson, 2003.

Gibbon, Edward. *The Decline and Fall of the Roman Empire*. Edited by David Widger. Salt Lake City, UT: Project Gutenberg, 1996.

Heather, Peter. *The Fall of the Roman Empire: A New History of Rome and the Barbarians*. New York: Oxford University Press, 2006.

Jones, A. H. M. *The Decline of the Ancient World*. London: Longman, 1966.

Jones, A. H. M. *The Later Roman Empire, 284–602: A Social, Economic and Administrative Survey*. London: Blackwell, 1964.

MacMullen, R. *Corruption and the Decline of Rome*. New Haven, CT: Yale University Press, 1989.

Mierow, Charles Christopher, trans. *The Gothic History of Jordanes*. English version with an Introduction and a Commentary. Princeton, NJ: Princeton University Press, 1915.

Nardo, Don. *The Decline and Fall of the Roman Empire*. World History Series, 27. San Diego: Lucent Books, 1998.

Potter, David S. *The Roman Empire at Bay. CE 180–395*. Routledge History of the Ancient World. London: Routledge, 2004.

Procopius. "The Vandalic War." In *The History of the Wars*, Books III and IV. Trans. by H. B. Dewing. Cambridge, MA: Harvard University Press, 1916.

Ward-Perkins, B. *The Fall of Rome and the End of Civilisation*. New York: Oxford University Press, 2005.

12

The Hawaiian and other Polynesian seafarers developed navigation methods based on observation of constellations and currents, so that they could sail intentionally from Tahiti to Hawaii and back.

PRO	Harald Haarmann
CON	Claire Brennan

PRO

The exploration and colonization of the islands in the Pacific Ocean have puzzled generations of scholars and have spurred wide-scale speculation among a broader public. Over a span of more than 200 years, the debate about how the Pacific was settled has produced an enormous amount of literature, ranging from observations about the lives and customs of Polynesian islanders in voyagers' logbooks to scientific investigations about ancient seafaring, archaeology, and population genetics.

No other region in the world offers so many similarities among local cultures and so many resemblances among languages that are genealogically related, scattered, as they are, over thousands of islands over an extremely vast area called Oceania. The geographic distances in Oceania are enormous. It is some 6,000 nautical miles (or 10,000 km, respectively) from Australia or New Guinea in the west to Easter Island in the east and some 5,000 nautical miles (about 8,500 km) from the Hawaiian archipelago in the north to the South Island of New Zealand.

Seafaring in the western Pacific and on the Pacific Rim started at a very early time. Those who crossed the strait that separated Southeast Asia from New Guinea during the Ice Age were the first seafaring humans. They undertook their voyages in canoes or on rafts more than 62,000 years ago. At that time, New Guinea and Australia were still interconnected in the land mass known as Sahul. Those early seafarers did not have any technical devices for their navigation, but they certainly observed phenomena in their natural surroundings very carefully and learned how to respond to the impact of environmental factors on their living conditions. Most probably, the people who crossed the waters to reach Sahul had some, albeit rudimentary, experience with currents and star constellations.

For a long time, reliable information about seafaring and colonization in the Pacific was scarce, a condition that kept spurring speculation, often revealing a

257

tendency to degrade the Polynesian ancestors' navigation skills and to underestimate their inventiveness. Most of this speculation originated at a time when the zeitgeist of colonialism favored a consciousness of superiority among Europeans and Americans, who saw those who did not work metal, did not possess writing, sailed without naval charts, and lived their lives without the blessings of the monopoly of a male god as primitive. Who would credit primitive islanders for sophisticated schemes of seafaring and specialized skills in navigation? It took well into the postcolonial era for the zeitgeist to shift from a marginalization of Polynesian achievements of seafaring to a genuine appreciation of their endeavors, untainted by old-fashioned misconceptions and worn-out prejudices.

Linked by Language

Those people who set out on their deep-sea voyages were members in sustained communities, united by the same language, rather than cast together by the vicissitudes of local conflicts and warfare. The history of the explorations of Oceania is the history of the eastward spread of Austronesian languages. The Austronesian family of languages (or phylum) is the second largest phylum (after the Niger-Congo language family) in the world, according to the number of genealogically related languages (i.e., some 1,230 individual languages). Most of the Austronesian languages belong to the Malayo-Polynesian branch. The cultural and linguistic unity of those Austronesians who colonized the Pacific islands is manifested in the fact that all languages of the Pacific form a homogeneous subgroup of Malayo-Polynesian: Oceanic (i.e., Hawaiian, Tahitian, Fidjian, Samoan, Maori, Rapanui, and other local varieties of Oceanic). Moreover, seafaring in the Pacific is just the eastern scenario of Austronesian migrations.

Impetus for Migration

Austronesian-speaking seafarers also sailed west and reached Madagascar off the coast of East Africa. The distance between that island and Southeast Asia is about 4,700 nautical miles (about 7,500 km). The true magnitude of the global endeavor of the Austronesian migrants can be seen in the extent of their western and eastern explorations. The overall picture of sea-bound movements of the Austronesians does not favor any speculation about their seafaring as being accidental.

If those who set sail for distant islands had been refugees and displaced people, leaving areas devastated by warfare between local clans, accounts about their rescue would have found their way into local myths of origin among the Polynesians. However, such sagas of alleged rescue operations are not typical of Polynesian mythology. Even if turmoil caused by warfare might have been the driving force for emigration in some cases, these are of rare occurrence and remain isolated in the long-term chronology of intentional and organized seafaring in Oceania.

For an understanding of the intentionality of the colonization movement in the Pacific region it is important to perceive the worldview of the Polynesians. This means that outsiders—whether Americans, Europeans, or people from Australia, Asia, or Africa—have to take into consideration the perception of Oceania in the cultural memory of the islanders. A modern observer who looks at a map of the Pacific region is impressed by the huge masses of water. This sort of vastness may evoke negative emotional associations such as emptiness of uncontrollable space, perils at sea, untraversable distances, vicissitudes of weather conditions, and the like. The mindset of a Polynesian, however, is differently tuned. An islander perceives the wholeness of Oceania in proportions different from outsiders. The many islands, scattered throughout the region, gain in profile as landmarks in a cultural landscape. The waters of the ocean are not perceived as unpredictable or even hostile. Instead, the ocean takes the role of a pathway connecting islands and communities, rather than functioning as a boundary to separate them. The vastness of the Pacific Ocean forms part of the islanders' cultural space, a space for interaction between nature and human beings. This mindset can be reconstructed from early maps that local guides drew for European sailors in the early days of contact. In these maps, the masses of water are drastically reduced, whereas the islands are given emphasis as spaces for human agency.

The Chronology of Migrations in the Pacific Region

The colonization of the Pacific Islands unfolded in a long-lasting process of several migratory waves, with intervals of stabilization of newly explored sea routes. It was mainly Austronesian-speaking people who explored and colonized Oceania. The migration of Austronesian populations begins with the second wave (starting around 1500 BCE). Those migrants profited from the experiences of seafaring of their predecessors who were not of the same ethnic stock as the Austronesians. The technology of building vessels apt for covering long nautical distances improved over generations, as did the methods of navigation, and it eventually enabled the Polynesians to explore the far-distant peripheries of Oceania (i.e., Easter Island, the Hawaiian Islands, and New Zealand).

The first migration (before ca. 40,000 BCE), coming from the mainland of Southeast Asia, to reach New Guinea more than 40,000 years ago were the ancestors of the modern Papuans, whose languages are unrelated to the Austronesian phylum. The first wave of migration proceeded as far as the Bismarck Archipelago and the Solomon Islands. The second migration (between 1500 and 1000 BCE) were Austronesian populations. It is still a matter of debate from where these people came: from the islands of Southeast Asia (west of New Guinea) or from their early settlements on the northern and eastern coasts of New Guinea. The Austronesian migrants reached the Carolines in the north, the Samoa group in the east, and New Caledonia in the south. The Fiji Islands are located in the central

part of this area (i.e., Melanesia) that was colonized during the second migration. The Melanesian cultural complex, centered on Fiji, provides the foundation for the subsequently developing Polynesian culture, which proliferated into a wide array of local cultures during later migrations. The proliferation of cultures found its parallel in the direction of gene flow. The Polynesian lineages stem, genetically, from the gene pool of populations in western Melanesia. The third migration (starting ca. 200 BCE), starting from the Fiji group, was directed to the east. The settlements on the Society Islands, with Tahiti as their cultural center, date to the time of the third migration. The fourth migration (starting ca. 300 CE) saw the greatest distances in the Pacific Ocean traversed by migrants from Tahiti. The distant Easter Island was settled around 300. The sea route to the north of Tahiti brought settlers to the Hawaiian Islands.

The fifth migration (10th century) involved the exploration of the sea route to the southwest of Tahiti, and migrants reached the islands of New Zealand toward the end of the 10th century. The last of the bigger islands to be settled was the Chatham group, east of New Zealand. The first Polynesians to reach Chatham around 1000 most probably arrived there, not from New Zealand, but on an independent voyage from Tahiti. The Moriori culture of Chatham declined in the first half of the 20th century.

The time frame for the more recent migrations are conservative estimates. The more calibrated radiocarbon dates become available, the more accurate the association with the absolute timescale will be. It has been assumed that early

War canoe used by the Maoris, the indigenous people of New Zealand, 19th century illustration. (Gianni Dagli Orti/Corbis)

settlers may already have reached the Marquesas Islands in far east Polynesia at the beginning of our era (that is by 0 CE), several hundred years earlier than according to the conservative time frame.

Migrations Planned

It has been demonstrated by means of computer simulations of the migratory movements in space and time that the peopling of the Pacific Islands cannot have occurred as the result of accidental drifts. Such accidental drifts would have produced random patterns, with densely populated islands contrasting with widely unpopulated areas. The only explanation for the regularities and the emergence of the dense web of local settlements is that the migrations were carefully planned. Polynesian demography is the outcome of intentional seafaring. In time, the seafaring endeavors became more frequent and better organized. This was not only the result of farsighted planning of voyages but also the accumulation of knowledge about climatic conditions and the advancement of technological development. Each wave of migrants could profit from the experiences of previous generations to improve their know-how of seafaring. Many of the sea routes that had been explored in the course of colonization were already regularly frequented for trading more than 2,000 years ago. The migrants from the second wave onward who set out on their voyages were agriculturalists, and they transferred their knowledge of food production (as horticulture and full-scale agriculture, where this was possible) to newly explored islands. To establish new settlements at a distance from the existing ones, it was necessary to transport plants and

Nainoa Thompson and the Science of Polynesian Navigation

One of the most persuasive arguments in favor of Polynesian navigational techniques is the fact that they have been re-created and used in the modern era. Starting in the 1970s, Nainoa Thompson began to study the ancient Polynesian science of way-finding for long-distance ocean voyaging, and has since then engaged in several journeys without the benefit of modern navigational equipment. In 1980 Thompson, his mentor Mau Piailug, and their crew made the trip from Hawaii to Tahiti aboard their voyaging canoe, the *Hokule'a*. They completed the over 2,500-mile journey through open ocean without the aid of any modern navigational devices. During 1985–1987, Thompson took the *Hokule'a* on what he called the "Voyage of Rediscovery," navigating all over Polynesia, visiting Tahiti, the Cook Islands, New Zealand, American Samoa, and 250-island atoll of Rangiroa before returning to Hawaii. During the early 1990s, Thompson built another voyaging canoe, named the *Hawai'iloa*, this time using all native Hawaiian materials, rather than the modern materials used to construct the *Hokule'a*. However, the *Hokule'a* had one more important voyage to make, this time completing the so-called Polynesian triangle by reaching Rapa Nui, or Easter Island, on October 8, 1999.

livestock (i.e., pigs, fowls, and dogs). To safeguard the continuity of new settle-ments over generations, the migrant groups had to include both men and women.

Long-Distance Voyage Navigation Knowledge

The Polynesians did not have any technical devices to facilitate their naviga-tion, and they did not possess systems of notation to draw charts of currents or of star configurations. What they definitely instrumentalized for their orienta-tion at sea was their refined perception of natural phenomena. Over time, a vast amount of specialized knowledge about seafaring accumulated, and this knowl-edge was transferred, by means of oral memory, from master to disciple in the professional domains of specialized handicraft. The builders of canoes and cata-marans profited from the experiences of seafarers to make their vessels apt for maneuvering against high waves and gusty winds. And the settlers who took livestock and provisions on their journeys learned from the stories that earlier migrants told to make improvements for their selection.

A crucial question that has puzzled many scholars is the proportional rela-tionship between one-way voyages (go and no return) and two-way voyages (go and return). Trade and cultural ties can only be kept up under the condition that voyagers explored new terrain for settlement and returned to their base to report about their discoveries. There are areas in Oceania that were settled in one-way voyages and remained isolated. This is true for Easter Island and for New Zealand. Geographic distance, though, was not a decisive condition for either isolation or permanent contact with other islands. The Hawaiian Islands are located on the extreme northern periphery of Oceania, thousands of miles away from other major island groups. Nevertheless, a trade route existed between Hawaii and Tahiti. An experienced seafarer would certainly explore the chances for a safe return, especially when sailing out into unknown space. One method of safeguarding return was sailing upwind. Crosswind sailing is more adventur-ous, but this kind of movement "became more sure when expanding geographi-cal knowledge allowed a return to land downwind of the point of origin" (Irwin 1992: 102). Computer simulations of far-distant voyages, undertaken by groups of migrants in several canoes, and their chances of two-way success show that the ratio for the loss of canoes and their voyagers was fairly low.

Three major factors used extensively by the Polynesians are essential for successful navigation under natural conditions: the knowledge of climatic con-ditions, experience dealing with currents, and the orientation at the position of major stars and star configurations in the nightly sky.

First, the knowledge of climatic conditions: In the tropical zone of Oceania, the air is heated up over the water masses and rises. This warm air drifts toward the north and south and cools down, and some of it streams back to the equator. Additional cold air streams down from the northern and southern latitudes,

collides with the hot air, and causes turbulences, eventually producing tropical storms. Storms that originate over the ocean are called *taifuns* (a Japanese expression) in the Pacific, cyclones in the Indian Ocean, and hurricanes in the Caribbean. The streaming of the air follows regular movements, creating elementary patterns for the direction of winds, so that "air flowing toward the pole gives winds from the west and air flowing toward the equator gives winds from the east" (Irwin 1992: 9). The polarization of west-easterly winds was one of the cornerstones of observation on which the Polynesian seafarers built their navigation system. Local conditions produced further variations of wind patterns. For instance, experience taught seafarers that the winds in January are the most favorable for sailing east from the Solomon Islands. The observance of winds and their direction is one thing; exploiting such winds for the purpose of practical seafaring is quite another. An astounding level of sophistication of craftsmanship is revealed in the Polynesian technology of constructing vessels apt for long-distance sea traveling and in the sailing equipment of those vessels. These vessels were not built by unskilled craftsmen, and they were not built for random travel. They reflect the accumulated know-how of many generations of seafaring people. The big canoes for seafaring were outrigged in Micronesia and double hulled in Polynesia, with a length of up to 22 meters. They could sail at 8 knots and cover up to 150 nautical miles in a day. The aptness of the canoes for seafaring is manifested in every detail of the technical equipment. The most effective type of sail in Polynesia is the sprit sail, a "triangular sail mounted apex downward" (Finney 1977: 1278). In such vessels, provisions could be stored from one to three months. In a month's voyage, these canoes could traverse distances of several thousands of nautical miles. It is noteworthy that, from the beginning of the migration movement, the winds posed challenges rather than played the role of pleasant forces that could be exploited by Polynesians intending to set out to sea. During the early migrations, the seafarers had to sail against the prevailing winds. Later, the movement was across and down the winds. The conditions changed again when the migrations reached beyond the tropics into the zone of temperate climate.

Second, the knowledge of oceanic currents: The great and stable currents were known to the early seafarers because they had to maneuver within or across them. Both the north equatorial current and the south equatorial current stream from the east in a westerly direction. The equatorial countercurrent streams from the west to the east, from the Carolines to Central America, reaching the coast in the region of Panama. In the eastern Pacific, the circular anticlockwise movement of the waters caused by the Humboldt (or Peru) current has a bearing on the direction of local currents as far as the Marquesas. Traveling from Hawaii to Tahiti was a demanding endeavor, and such a journey required crossing easterly winds almost constantly. The easiest part was the final passage from Tuamotu to Tahiti. Sailing from Tahiti back to Hawaii might have been easier, with southeasterly winds taking vessels as far as beyond the equator.

Finally, the knowledge of star positions and of star configurations: Several factors make the observation of the nightly sky in the Pacific region more favorable and practical for orientation than in many other regions of the world. One of these factors is the weather conditions. Statistically, the tropical and subtropical zones offer more sunshine than the temperate zone, with its higher probability for cloudy weather. Comparatively, these favorable conditions made the observation of stars and their configurations a more practical means of orientation for Polynesian seafaring than for the Vikings in the Northern Hemisphere. Another factor is that air pollution caused by human agency is minimal in Oceania, so that different categories of brightness of stars can be discerned with the naked eye. The air is so clean that the contours of the Milky Way are visible. Consequently, star configurations gain in profile under these conditions.

Astronomers can tell from their experience that the Southern Hemisphere is "richer" in stars and star configurations than the Northern Hemisphere. Since a greater variety of star categories is visible in the clean air of Oceania, this also means that stars that are not bright are visible, providing a dim-light background for the bright stars. The verified constellations of stars in the Pacific region, however, revealed themselves to western seafarers as late as the 16th century. Before then the Europeans used stellar charts from antiquity. The best known of these are the compilations of charts in the work "Megale Syntax" (Great Syntax) of the astronomer Ptolemaeus (second century CE). The original Greek version got lost, but its contents were preserved in an Arabic translation, which, in turn, were retranslated into Latin in 1175.

The Polynesian seafarers were well acquainted with the star configurations of the Southern Hemisphere. Practically every major group of islands in Oceania has its own particular star configuration. Using the position of single bright stars and these configurations for navigation was not as easy as looking at the sky, because the constellations, forming sections in the bigger mosaic of visible heavenly bodies, may shift their relative position depending on the seasons. At times, certain configurations are invisible. The seafarers had to learn from the experience of earlier generations what stars and configurations to orient to where and when.

Navigators would rely on the observation "from the rising point and trajectory of a large number of familiar stars—steering by 'star paths' enables the skilled navigator to make allowances for the effects of drift by winds and currents" (Nile and Clerk 1996: 63). The knowledge of particular star configurations and their position in association with certain islands assisted seafarers in choosing and keeping direction to reach the intended goals of their voyages. Orientation at the stars alone would certainly not have sufficed for successful navigation. The navigation skills of the Polynesians were anchored in the interplay of all variables of seafaring, with the identification of familiar stars forming part of a web of orientation, also including the equally important observance of and interaction with currents and winds.

Importance of Regular Seafaring and Trading

Once sea routes had been explored, these were used not only for the transport of settlers and their livestock but also for the transfer of trade goods. The earliest evidence for overseas trading activities comes from Melanesia. The most important merchandise to be traded in early times was obsidian as raw material. In the course of time, a greater variety of goods were traded. Since sea routes for trading had already been explored in Melanesia before the colonization of the Polynesian islands, the history of seaborne trading is older than the Polynesian settlements, and the Melanesian tradition was inherited by the Polynesians.

The evidence documenting the existence of ancient trade routes—for short-distance as well as for long-distance trading—speaks in favor of organized seafaring. Seafaring in terms of frequenting certain sea routes in regular intervals without navigational skills is unthinkable. The simple opening of a trade route requires knowledge about the conditions of two-way voyaging. In concrete terms, this means that the traders have to know that there is a place out at sea where they will find buyers for their goods, and they also have to know how to get there and how to return. Seaborne trading grew in magnitude and importance for the interconnection of scattered settlements in Melanesia and throughout Polynesia.

According to the time depth of their appearance, a number of goods played a major role in sea trading. Obsidian is volcanic glass that was used in many parts of the world for tools. Obsidian flakes provide sharp cutting edges. This raw material was found at two sites on the island of New Britain in the Bismarck Archipelago off New Guinea. One is Misisil, where obsidian might have been obtained by the pre-Austronesian islanders more than 10,000 years ago. The other site is Talasea. Already around 4500 BCE, Talasea obsidian was transported to the neighboring island of New Ireland, at a distance of some 30 km from New Britain. In our era, Talasea obsidian found its way as far as New Caledonia, Vanuatu, and Fiji. Obsidian was not an isolated article for trade, but it often appeared together with *lapita* pottery, which is tempered with sand and hardened in open fires. It has produced a variety of forms such as cooking pots in a globular shape, bowls, and dishes with a flat bottom. The *lapita* pots were often decorated with geometric motifs, sometimes also with stylized naturalistic pictures of human beings. The earliest evidence for *lapita* production is known from the Bismarck Archipelago, dating to ca. 1600 BCE.

By the time *lapita* production declined (around 500 BCE), this type of pottery had spread throughout Melanesia, far beyond the area of settlements where the workshops for these goods were located. The evidence thus suggests "that early Lapita history was not one of isolated settler communities but, rather, involved continuing contacts and exchanges based on two-way voyaging in sailing canoes" (Nile and Clerk 1996: 55). The *lapita* voyagers traversed longer distances than the early obsidian traders (more than 1,600 miles, or 2,500 km).

Basalt (as raw material): Because of its qualities, this special stone was the preferred raw material for adzes. The earliest adzes were made of shells and later of local stone. The Tuamotu group of islands (to the northeast of Tahiti) offered a special variety of basalt that was traded over long distances. Since Tuamotu lies on the sea route from Hawaii to Tahiti, it is not surprising that Tuamotu basalt as a raw material for adzes is found both in Hawaii and in Tahiti. The distance between the two island groups is about 3,000 nautical miles (about 4,800 km).

A modern reconstruction of the Hawaiian two-sailed canoe, called *Hokule'a* in the Hawaiian language, was used in 1976 to make a voyage from Maui in the Hawaiian Islands to Tahiti. This voyage was made without the help of modern instruments of navigation, with star configurations and the observation of currents and winds as the only means of orientation. The voyage was successful and may serve as a demonstration to illustrate the possibilities of seafaring over long distances in historical times. Another voyage with the same vessel was undertaken from central Polynesia to New Zealand several years later, and this endeavor was also successfully concluded.

Ceremonial Voyages

Voyaging was not only regularly practiced for the transport of migrants, livestock, and trading goods but also for revitalizing social contacts between settlers and for reaffirming cultural relationships between settlements that were separated by the waters of the ocean. Traditions to conduct ceremonial voyages originated in many parts of Oceania. Perhaps the network of ceremonial voyages with the widest geographic range is the *kula* network of gift exchange in Melanesia. To perform the ceremonial voyages in connection with this gift exchange, relatively long distances have to be traversed in special ceremonial canoes. The settlements that participate in the *kula* network cover an area of some 250 nautical miles (about 360 km) in diameter, from the southeastern tip of New Guinea in the west, the Laughlan Islands in the east, the Trobriand Islands in the north, and the Louisiade Archipelago in the south.

The annual voyages of the *kula* gift exchange are performed in richly decorated outrigger canoes. The decorations are painted and worked with cowrie shells. The men in the village communities organize themselves in groups to visit their trading partners overseas. The ritual exchange of gifts includes shell necklaces (*soulava*), which move in a clockwise direction, and armshells (*mwali*), which move anticlockwise.

References and Further Reading

Conte, E. "L'homme et la mer." *Encyclopédie de la Polynésie* (Papeete, Tahiti), Vol. 5 (1990): 41–56 (including a series of illustrations of canoe types and sailing gear).

Finney, Ben R. "Voyaging Canoes and the Settlement of Polynesia." *Science* 196, 4296 (1977): 1277–85.

Garanger, J. "Le peuplement de la Polynésie orientale." *Encyclopédie de la Polynésie* (Papeete, Tahiti), Vol. 4 (1990): 41–56.

Irwin, Geoffrey. *The Prehistoric Exploration and Colonization of the Pacific*. Cambridge: Cambridge University Press, 1992.

King, Michael. *Moriori—A People Rediscovered*, 2nd ed. Auckland: Viking, 2000.

Kirch, P. V., and R. C. Green. *Hawaiki, Ancestral Polynesia: An Essay in Historical Anthropology*. Cambridge: Cambridge University Press, 2001.

Nile, Richard, and Christian Clerk. *Cultural Atlas of Australia, New Zealand and the South Pacific*. Abingdon: Andromeda Oxford Limited, 1996.

Oppenheimer, Stephen, and Martin Richards. "Polynesians: Devolved Taiwanese Farmers or Wallacean Maritime Traders with Fishing, Foraging, and Horticultural Skills?" In *Examining the Farming/Language Dispersal Hypothesis* (pp. 287–97). Edited by Peter Bellwood and Colin Renfrew. Cambridge: McDonald Institute for Archaeological Research, 2002.

Thornton, Agathe. *The Birth of the Universe—Te whanautanga o te ao tukupu*. Auckland: Reed Books, 2004.

CON

The debate over Polynesian migrations is a long running one. Early European voyagers in the Pacific were perplexed by the existence of people who were obviously culturally related but who inhabited the widely dispersed islands of the Pacific. European navigators were particularly confused as their own technology had only just allowed them to voyage to these islands, and yet on landing they seemed to be inevitably confronted by people using stone-age technology who had reached these small and widely scattered pieces of land well before them. Once appeals to divine intervention were no longer thought a sufficient explanation for Polynesian dispersal, speculation about their methods of finding and settling islands became widespread, and among the explanations proposed was that of accidental voyaging.

The area bounded by Tokyo, Jakarta, and Easter Island contains 80 percent of the world's islands. The islands of the Pacific region are generally grouped into the geographic regions of Melanesia, Micronesia, and Polynesia, and debates about navigation focus on Polynesia. The islands of Polynesia are generally described as being contained within a triangle, with the corners comprised of New Zealand, Hawaii, and Easter Island. The corners of the triangle are distant from everywhere. Hawaii is approximately 2,000 miles from the North American continent (it is about equally distant from California and the Aleutian Islands), and 600 miles

from Johnson Atoll, a small piece of land. It is 2,400 miles from Tahiti, and 3,800 miles from New Zealand. Similarly New Zealand is about 1,300 miles distant from its nearest continent (Australia) and only slightly closer to Fiji, the nearest island group. Easter Island is often described as the most isolated place on Earth, as it is 2,400 miles west of South America and 1,200 miles east of Pitcairn, the nearest land. The region of central Polynesia contains islands much closer together; it is travel to these distant corners of Polynesia that is involved in this debate.

The debate over the navigational practices of Polynesians continues into the present because the existence of a recognizably single culture across this area of the world is an astonishing feat. The Pacific is vast—it is greater in area than all the world's land masses combined, and larger even than the rest of the world's oceans combined. The difficulty of settling Polynesia meant that it was the last region on Earth to receive humans. While Melanesia could be settled by a process of island hopping—moving from one island to another always within sight of land—settlement of Polynesia required new techniques as voyagers had to set out without being able to see their destination, and travel without land in sight for long periods of time. Melanesia was fully settled approximately 3,200 years ago, but Easter Island was only reached about 1,700 years ago, Hawaii 1,600 years ago, and people arrived in New Zealand possibly as recently as 700 years ago. Whether settlement of this region involved deliberate voyaging or contained elements of chance remains an open question.

The idea of regular deliberate voyaging between the far reaches of Polynesia has its problems. Certainly such voyages were not occurring when Europeans entered the Pacific, and the possible place of chance in Polynesian settlement should not be ignored. While the idea of Polynesian navigation across the vast Pacific is presently fashionable, older ideas about the role of drift voyaging in Pacific settlement deserve consideration. Before examining support for the theory of drift voyaging, it is worth noting what it is not: the idea that Polynesians discovered their islands by chance does not deny that Polynesians, when encountered by Europeans, used canoes and undertook ocean voyages. Instead the argument is over whether long-distance navigation—not observed by Europeans on contact—had occurred at some point in the past, and whether Polynesian people had deliberately voyaged between far-flung islands and had deliberately set off in quest of unknown land. Alternatively the drift voyage thesis maintains that the Pacific is a difficult environment and that some Polynesian settlement can be explained by voyages of chance when fishing parties or groups of refugees stumbled across islands when blown there by unexpected winds or pushed there by unknown currents.

Early Support for Drift Voyaging

As Europeans explored the Pacific they were fascinated by the people who lived there, recording observations of them and information gained from talking to

them. Part of early European interest in Polynesian navigation was pragmatic—Europeans wanted help in navigating the vast Pacific, and they were eager for information that could be gained from Polynesians, Tahitians in particular. Thus the famous British navigator Captain Cook eagerly compared information about the location of Pacific islands with his passenger, the Tahitian navigator Tupaia, who joined Cook's *Endeavour* voyage in 1769. While Tupaia was able to add 80 islands to Cook's charts, he did not impart any information about lands as distant as New Zealand or Hawaii, indicating that voyaging between Tahiti and these places did not occur at that time and had not occurred for a long time, if at all. Similarly, when the Spaniard Andia y Varela visited Tahiti in 1774–1775 he, too, picked up a Tahitian navigator. Andia recorded the ability of the Tahitian to set courses to travel to islands that were not visible from Tahiti itself, but the islands known in this way were limited to those not too far distant, and such islands were in regular contact with Tahiti. The far corners of Polynesia were not known to Tahitians at the time of contact with Europeans, showing that long-distance voyaging was not occurring and had not occurred in the recent past.

This experience of a lack of knowledge of the far corners of Polynesia was not limited to Tahiti, but occurred throughout the islands of central Polynesia. While in Tonga, Cook collected information about other Pacific islands from his hosts. That list included 156 islands known to his Tongan informants, but despite this wealth of knowledge, islands that required open-water journeys of more than 30 miles from Tonga were not included. Thus large and significant island groups in the vicinity but more than 30 miles distant (such as the Cook Islands, Tahiti, and Niue) were not among those described to Cook. It would seem that Tongan navigation covered only limited legs of the ocean, although island hopping along chains of islands meant that Tongans were aware of many islands. However, the information collected by Cook indicated no knowledge in Tonga of the existence of Hawaii, Easter Island, or New Zealand, and again clearly indicated that regular voyages to any of those places did not occur at the time of contact and had not occurred in the recent past.

In this early period of contact there is no evidence that Tahitian navigators visited, or were even aware of, islands as far distant as Hawaii. Deliberate navigation took place, but it occurred only within limited regions, such as within the island groups of Tahiti, Hawaii, and Tonga/Samoa. Travel between these discrete groups, and between island Polynesia and New Zealand, was not observed and seemed not to have occurred at all recently. Tupaia was unaware of the location of Hawaii, and although Tupaia traveled with Cook on his first voyage, Cook did not visit the Hawaiian Islands until his third voyage to the Pacific. In Cook's observations the only indication of the existence of more distant islands were Polynesian stories of distant origins, but no voyages to the corners of the Polynesian triangle occurred. The mystery of Polynesian dispersal could not be easily solved.

As a result of the lack of any clear evidence of Polynesian long-distance navigation at contact, the first theorist to suggest that Polynesian navigators reached their islands by drift voyages was Cook himself. Cook regularly observed Polynesian canoes during his visits to island Polynesia, and he was impressed by the size, speed, and maneuverability of those that he observed in Tahiti and Tonga. However, even after extended contact with Tupaia, Cook argued that Tahitian navigation did not deal with very long voyages and did not have the tools to cope with long periods at sea out of sight of land. In addition to his observations of the limits of Tupaia's knowledge of Polynesia, Cook also observed evidence of drift voyages occurring across long distances. While in the southern Cook Islands he observed survivors of a drift voyage from Tahiti. A canoe had been blown off course, and although 15 members of the crew had died, five had survived and landed in Tahiti. Cook argued that such chance events accounted for the settlement of Polynesia, as people in canoes were blown away from known land, and some were fortunate enough to land on previously unknown islands and establish societies. No evidence of more deliberate long-range voyaging was found by Cook.

Postcontact Changes in Polynesian Traditions

The stories of distant origins noted by Cook at contact have led to speculation about repeated long-distance voyages, but such sources must be used with care. At the time of Cook's first voyage, closer connections were developing between the Pacific and Europe, muddying the waters of the Pacific as far as the extent of traditional navigational knowledge was concerned. Polynesians were eager to incorporate European geographic knowledge into their own, and Europeans were fascinated by Polynesian peoples and set about collecting their traditions. The transmission of Polynesian traditions into European language and culture, and then back again, inevitably altered them, and "pure" Polynesian traditions have been lost as traditions evolved in the face of new ideas. For example, Cook's use of Maori names for the islands of New Zealand, not available to Tahitians before his voyage there, were given to Andia in 1774–1775 as part of Tahitian navigational knowledge. The Maori names for the islands of New Zealand were not part of the traditional navigational knowledge of Tupaia. However, when Cook discussed his voyages with Tahitian navigators, he introduced them to the existence of New Zealand, using those names. That knowledge was adsorbed and quickly incorporated, and the names for New Zealand's islands became considered traditional knowledge and was not differentiated as originating from a later period than other navigational knowledge.

European fascination with Polynesian tradition only increased as Europeans came to settle in New Zealand and Hawaii. The first theorist to propose large-scale,

Captain James Cook Describes a Polynesian Canoe

The first European to voyage through Polynesia was the famous British captain James Cook. His writings formed the first notions the rest of the world would have of the Pacific as a whole and of its people. He wrote at length of Polynesian seafaring, and what follows is his description of the canoes the native people used.

The ingenuity of these people appears in nothing more than in their canoes: they are long and narrow, and in shape very much resemble a New England whale-boat: the larger sort seem to be built chiefly for war, and will carry from forty to eighty, or a hundred armed men. We measured one which lay ashore at Tolaga: she was sixty-eight feet and a half long, five feet broad, and three feet and a half deep; the bottom was sharp, with straight sides like a wedge, and consisted of three lengths, hollowed out to about two inches, or an inch and a half thick, and well fastened together with strong plaiting: each side consisted of one entire plank, sixty-three feet long, ten or twelve inches broad, and about an inch and a quarter thick, and these were fitted and lashed to the bottom part with great dexterity and strength. A considerable number of thwarts were laid from gunwale to gunwale, to which they were securely lashed on each side, as a strengthening to the boat. The ornament at the head projected five or six feet beyond the body, and was about four feet and a half high; the ornament at the stern was fixed upon that end, as the stern post of a ship is upon her keel, and was about fourteen feet high, two feet broad, and an inch and a half thick. They both consisted of boards of carved work, of which the design was much better than the execution. All their canoes, except a few at Opoorage or Mercury Bay, which were of one piece, and hollowed by fire, are built after this plan, and few are less than twenty feet long: some of the smaller sort have outriggers, and sometimes two of them are joined together, but this is not common. The carving upon the stern and head ornaments of the inferior boats, which seemed to be intended wholly for fishing, consists of the figure of a man, with a face as ugly as can be conceived, and a monstrous tongue thrust out of the mouth, with the white shells of sea-ears stuck in for the eyes. But the canoes of the superior kind, which seem to be their men-of-war, are magnificently adorned with open work, and covered with loose fringes of black feathers, which had a most elegant appearance: the gunwale boards were also frequently carved in a grotesque taste, and adorned with tufts of white feathers placed upon a black ground. Of visible objects that are wholly new, no verbal description can convey a just idea, but in proportion as they resemble some that are already known, to which the mind of the reader must be referred: the carving of these people being of a singular kind, and not in the likeness of anything that is known on our side of the ocean, either "in the heaven above, or in the earth beneath, or in the waters that are under the earth," I must refer wholly to the representations which will be found of it in the cut.

Source: James Cook. *The Three Voyages of Captain James Cook*. London: William Smith, 1842.

long-distance deliberate voyages by Polynesian navigators in the precontact period was Percy Smith, who was a keen early 20th-century collector of Maori traditions. However, Smith is now acknowledged to have fundamentally altered Maori traditions as he collected and recorded them, tidying and reinterpreting stories collected earlier to fit in with his expectation of a unified mythology. In an attempt to rationalize the stories available from Maori and earlier European sources and to create a coherent framework of tradition from which he could extract further information, Smith in effect grafted European myths about Maori onto Maori mythology. Many of his ideas about the arrival of Polynesians in New Zealand, including the figure of Kupe (a great navigator who made a return voyage between tropical Polynesia and New Zealand in ca. 750 CE) and the idea of the Great Fleet of seven canoes that arrived and settled New Zealand, were adopted by Maori themselves and have become accepted as prehistoric Maori traditions. These ideas were exposed as European rather than Maori myths by D. R. Simmons in 1969, and his work has subsequently been consolidated but not challenged, yet the myths involved have not been abandoned. While raising interesting questions about the extent of Polynesian voyaging, the work of collectors of tradition produced altered versions of Maori tradition and in the process destroyed the original. This means that stories such as those of the Great Fleet cannot be understood as unadulterated Maori tradition or as a record of previous periods of deliberate long-distance navigation.

Similar traditions were found or invented in Hawaii by Abraham Fornander (1878). In a manner very similar to the collectors of myths in New Zealand, Fornander was fascinated by the mystery of Polynesian origins, using an unsophisticated linguistic analysis to (wrongly) give them Aryan origins, with a further connection to the Mediterranean. Among the stories that he collected, and in writing down and editing unsuspectingly changed, was that of the great navigator Hawai'iloa who traveled from his homeland to Hawaii. Hawai'iloa was reported by Fornander to have returned home to collect his family, and then made a second successful landfall in the Hawaiian islands. Although 19th- and early-20th-century anthropologists might consider such stories to provide evidence of regular long-distance Polynesian voyaging, too many questions have been raised about the process of recording and then reremembering what was involved in their codification for weight to be placed on them as evidence of repeated long-distance voyaging.

Andrew Sharp and Drift Voyaging as an Explanation for Polynesian Distribution

While these adulterated traditions were seen as providing evidence of long-distance Polynesian navigation, questions about the feasibility of such voyages continued to emerge. While not necessarily promoting purely accidental voyaging, some commentators did question tales of repeated voyages between such

distant points as Tahiti and Hawaii. For example, in 1924 John Bollons, an experienced sailor, wrote:

> It is amazing how boldly the landsman has launched the Polynesian out into the—at that time—infinite ocean; described the voyage, the seaworthiness of the canoes, the cargo carried, the manner in which the canoes were hove-to in bad weather, and the navigating by stars and the rising and setting of the sun. How simple it is, or seems to be, when one is living ashore" (quoted in Bellwood 1979: 301).

The distances proposed for deliberate and repeated navigation were immense, and Bollons was not the only sailor to point out that navigating the Pacific with modern tools was not always simple, and that doing so without the benefit of modern ship-building technology or navigational instruments was not as simple a task as it might seem to those who had never been in a boat out of sight of land with the sky clouded over and a shifting wind blowing.

Yet it was a landsman, Andrew Sharp, who launched the fiercest attack on the romantic myth of Polynesian navigation. In 1956 Sharp published the book *Ancient Voyagers in the Pacific* and challenged the notion that ancient Polynesians navigated the vast expanse of the Pacific easily and often. Sharp did not argue that all long-distance voyages in Polynesia were solely the result of aimless drifting, but he did argue that the notion of repeated navigations between the center and far corners of Polynesia was not clearly supported by tangible evidence. Sharp's attack on the romantic notion of a vast Pacific highway did not deny the possibility of Polynesian navigation in all its forms, but rather it questioned the regularity and control that Polynesian navigators could exercise over voyages more than about 300 miles in length out of sight of land—especially those thought to connect the corners of the Polynesian triangle to the center. While Sharp argued for "drift" voyages, his definition of drift is problematic. His book argued that when a destination is unknown, no course can be set for it—that navigation to an unknown place is by definition impossible. This is not quite the same thing as drifting directionless across the ocean, but it was enough to ignite fierce opposition, as it seemed to challenge ideas about Polynesian cultural development and to strike at a source of pride for indigenous Polynesians.

Despite the opposition that Sharp aroused, his book raised useful (and at the time unanswerable) questions about the nature of Polynesian navigation. Responses to his work were immediate and heated, and some involved projects to undermine his argument by establishing what traditional navigation entailed. Certainly early enthusiasts for traditional navigation did not specify what tools the Polynesian navigator had available, nor how it was possible to steer a reliable and repeatable course without a compass, out of sight of land or how to cope with the effects of currents whose influence would be essentially invisible. Essentially very little information was available on how to navigate sufficiently precisely to locate small islands in a vast ocean when failing to find them would

mean dying at sea. Certainly such knowledge was not generally available in the middle of the 20th century, and it was Sharp's work that prompted attempts to save and revive such knowledge where it still existed. Before his publication, ideas of Polynesian navigation were based on vague ideas of using the position of the sun and stars without any clear indication of how this was to be achieved.

Sharp raised other useful questions that challenged unexamined assumptions about Polynesian long-distance navigation. He argued that the materials available to islanders when building their craft could not construct a craft capable of reliably sailing long distances. In particular, Sharp argued that the rope materials available within Polynesia were vulnerable to strain on long voyages and could not be relied on in rough conditions or when sailing against the wind. He also pointed to the seasonality of winds, and of stars, as guides to Pacific navigation. He noted that language differences had arisen within Polynesia, indicating a failure of continued communication. Similarly the incomplete distribution of Polynesian livestock (the dog, rat, pig, and chicken) throughout Polynesia argued against continuing, regular contact. Sharp also questioned the assumption that Polynesian canoe-building and navigational technology had declined by the time of Cook's first visit, arguing against the assumption that knowledge had previously existed and somehow been lost. And Sharp was able to produce evidence of well-documented cases of unintentional and unnavigated voyages across the Pacific in historic times. His book raised useful questions about the difficulty of long-distance navigation across the Pacific and about the role of drift voyages in the discovery of land, but they came under immediate attack because they contradicted those mythical ideas about Maori navigation and the settlement of New Zealand.

Sharp received some support for his ideas, notably from Kjell Akerblom in a 1968 Swedish publication. Such support was perhaps surprising as Sharp also challenged the notion that the Vikings had reached North America, considering the sagas provided unreliable evidence. (Subsequent archaeological discoveries have provided conclusive evidence for a short-lived Viking presence in North America.) Certainly Sharp's skepticism about Polynesian navigation was justified in terms of the material available to him. In response to his book much work was done on the ways in which Stone Age Pacific peoples might have navigated, including practical experiments recording the ways in which modern navigators still using traditional methods found their way between known groups of islands. It is worth noting that such material was mostly collected from Melanesia and Micronesia, as it had been lost in Polynesia when new navigational tools became available.

Thor Heyerdahl and the Implications of a South American Connection

However improbable, the most famous intervention in the argument about deliberate voyaging in the settlement of the Pacific involved a Norwegian. In 1947
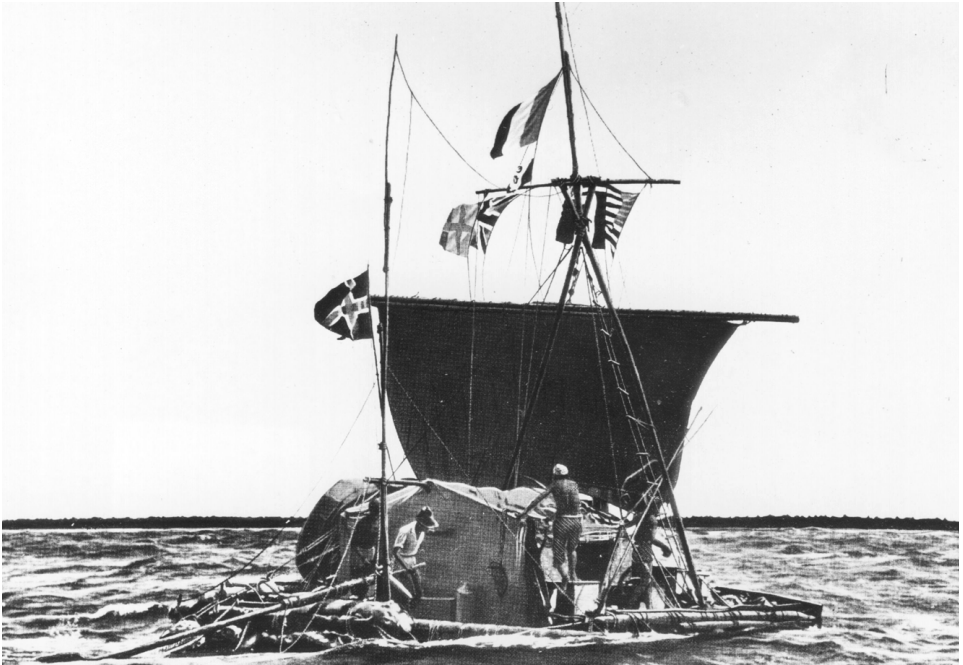
Norwegian ethnologist Thor Heyerdahl and his balsa raft *Kon-Tiki* crossing the Pacific Ocean on his drifting expedition from Peru to Polynesia, 1947. (Keystone/Getty Images)

Thor Heyerdahl set off across the Pacific on the balsa-wood raft, the *Kon-Tiki*. The raft was towed about 50 miles offshore from Peru, and then drifted and sailed across the Pacific before reaching the Tuamotu Archipelago, a distance of approximately 4,300 miles in 101 days. The voyage seized the public's imagination and demonstrated that despite the vast size of the Pacific, it was possible to run into land by chance rather than design. Other raft navigators of the Pacific followed in Heyerdahl's wake, not always successfully. Eric de Bisschop died in the Cook Islands in 1956 after a long-distance rafting accident, and in 1974 the crew of a junk expedition was rescued after five months at sea far from land in the northern Pacific. Heyerdahl's expedition was intended to deal with questions of Polynesian origins rather than their means of migration, and it drew on the older idea that because the prevailing winds in the Pacific blow from the east, people were blown into the Pacific from South America. Heyerdahl's South American origin for Polynesians has generally been rejected—on the grounds of language, a cultural trail of artifacts on the western edge of island Polynesia pointing to island hopping and settlement, and on physical characteristics—but his expedition indicated that long-distance voyages guided by luck rather than design could succeed and so produced evidence to support the possibility of significant drift voyages in the settlement of the Pacific.

Heyerdahl was not the only scholar interested in South American influences in the Pacific. Robert Langdon (2009) spent his career pointing out that the idea

of Asian origins cannot account for the spread of plants and animals throughout the Polynesian islands, as a number of significant species are of South American origin. The South American origin of species such as the *kumara* (sweet potato) cannot be denied, and the presence of such species again points to the possibility of drift voyages by South American rafts. Langdon argued that rafts that became dismasted were likely to drift west across the Pacific, and pointed to historical examples of such voyages, including that of Heyerdahl. He added the idea that weather patterns in the Pacific are not constant, and that El Niño weather patterns might have facilitated such drift voyages from South America, again citing voyages within recorded history, this time of two European vessels. The first was the voyage of an English pirate in 1687 who drifted from the coast of South America to an island that Langdon identifies as Easter Island. The second, more convincing, example was that of the HMS *Chanticleer* in 1870, which was carried into the Polynesian triangle while attempting to navigate between Panama and Callao. It stopped at Easter Island on its way back to the South American coast. Thus, evidence of South American influences within the Pacific has links to the question of the role of drift voyages in establishing Polynesian culture.

The Problem of Evidence

The major difficulty in this debate is the lack of solid evidence for either position. By the time written records began in the Pacific, with the arrival of Europeans in the 16th century, any period of navigated voyages to the corners of the Polynesian triangle was past. Instead of Polynesian canoes navigating the far reaches of the Pacific using traditional methods, Polynesian people quickly began traveling on European ships and adopting European ship technology and navigational instruments. In particular, Hawaiian men traveled the world on European and American whaling ships and also participated in inland fur expeditions in North America. In 1834 a federation of chiefs from the northern regions of New Zealand registered a flag in order to identify ships originating in New Zealand. Such a flag was useful to them because they owned and operated a fleet of European-style trading ships. This quick adoption of new materials and their associated technologies meant that the building of large traditional canoes was quickly abandoned throughout Polynesia, and information that might shed light on the abilities of those canoes was lost with them.

In terms of the larger question of Polynesian origins, the pathway of Polynesian migrations tends to be reconstructed through linguistic analysis and archaeology of pottery and stone tools, rather than direct evidence of surviving canoes. This is because the Pacific is a difficult environment for wood to survive in. Canoes in particular were faced with the rot inherent in wet surroundings and also with various burrowing creatures such as the Teredo worm. As a result, other than images made by Europeans at the time of contact, no direct evidence

remains of canoes even at that time. No direct evidence is available, or can reasonably be expected to have survived, from earlier periods. As a result, ideas about possible navigated long-distance voyages are based on speculation, not material evidence.

Reconstructed Voyages and Their Connections with the Present

The reemergence of indigenous cultural identity in the Pacific has clouded the issue further. Islanders who had been colonized by European powers were rediscovering their political identities in the 20th century. As a result, reaction to Andrew Sharp's book was in part a reaction to what was seen as an insult to the ingenuity and navigational prowess of their Polynesian ancestors. An early example of reemerging pride in Pacific traditions can be found in the work of Te Rangi Hiroa (Sir Peter Buck), an anthropologist of Maori descent who worked as director of the Bishop Museum in Hawaii. As early as 1938, his book *Vikings of the Sunrise* was a clear statement of pride in ancestral achievements, including the navigation of the Pacific. Further cultural renaissance followed in New Zealand, and later throughout the Pacific, and this renaissance tended to draw on ideas of deliberate voyaging across the vast expanses of the Pacific.

The cultural renaissance in Hawai'-i that began in the 1970s was clearly associated with experimental canoe reconstructions and voyages. The establishment of the Polynesian Voyaging Society and the management of its canoe *Hokule'a* were largely concerned with the contemporary cultural revival of native Hawaiians. That canoe was reconstructed using historical images and modern materials and in 1976 successfully navigated the 2,250 miles between Hawaii and Tahiti using reconstructed navigational techniques. Between 1985 and 1987 the *Hokule'a* navigated a 12,000 mile course between Hawaii and New Zealand. However, a Hawaiian attempt in the early 1990s to construct a large voyaging canoe using only traditional materials was forced to compromise for reasons of scarcity of resources and time and for reasons of safety, as traditional sail and rope materials could not be processed in a way to make them work for such a large canoe. That project did lead to the voyage of a fleet of reconstructed canoes between Rarotonga in the Cook Islands and Hawaii in 1992. While the voyage of a fleet of reconstructed canoes was a remarkable achievement, it was a remarkable achievement in the present, rather than direct evidence of deliberate voyaging in the past. The use of reconstructed methods of navigation and of reconstructed canoe styles made an important statement about the value of Polynesian cultural traditions, but the voyage also exposed the failings of some of the vessels that attempted to join the fleet and the difficulties of navigating the Pacific. A Tahitian-built canoe was found to be unseaworthy and could not participate, and at various times the requirements of safety and of timetables meant that the canoes had to be towed by modern vessels. The voyage established both that Polynesian vessels and navigational

techniques were well developed and worthy of pride, and that the Pacific is a difficult ocean that is not easily navigated and where winds are fickle, and the danger of capsize and destruction is prominent.

Recorded Drift Voyages

Such rediscovering of traditional navigation and voyaging methods was often an explicit response to the criticism felt to emanate from Sharp's work. However, while deliberate voyages in reconstructed canoes have been undertaken in recent years, accidental drift voyages have also occurred. These voyages have been unplanned, yet at times have covered extraordinary distances. Such an example was reported in 2001 by the BBC when a Samoan fishing boat was pushed out to sea by currents after losing the use of its motor. Of the four crew, two died during the journey, but two survived a 2,800-mile, 132-day voyage to Papua New Guinea. Similar long-distance drift voyages have been noted by all participants in the debate on Polynesian navigation, beginning with Cook and including numerous examples cited by Sharp. Drift voyages certainly occur within the Pacific, and some fortunately end in landfall. Whether this process could constitute a complete explanation for Pacific settlement is a separate question—that long-distance drift voyages have occurred and continue to occur is not in doubt.

Conclusion

The debate about navigation in the Pacific cannot be fully resolved. The evidence available is not conclusive—the watery environment means that physical remains of great voyaging canoes are unlikely to have survived to the present, and in the time since the settlement of the Pacific, particularly the past 200 years, remembered evidence has been distorted. In those past two centuries massive change occurred as a result of the availability of iron and other aspects of European material culture, and the ways in which Polynesians traveled underwent a fundamental transformation. Traditions have also been distorted in the process of collection and recording, making stories of repeated voyaging suspect. Current theories of the way in which Polynesians settled the Pacific and of long-distance navigation tend to reflect contemporary concerns. At present it is unfashionable to argue that drift played a prominent part in Polynesian settlement of the Pacific. However, despite the rediscovered skills of Pacific navigators and the impressive voyages of reconstructed canoes, the Pacific remains a difficult environment. The continued process of drift voyages across seemingly impossible distances cannot be dismissed, and the role of accident and drift in the settlement of Polynesia requires consideration. Polynesians were fine mariners and efficient navigators within island groups, but the Pacific is a vast ocean, and Polynesian canoes were certainly at times directed by wind and current rather than human will.

References and Further Reading

Akerblom, K. *Astronomy and Navigation in Polynesia and Micronesia*. Stockholm: Ethnographical Museum Monograph 4, 1968.

Belich, James. *Making Peoples: A History of the New Zealanders from Polynesian Settlement to the End of the Nineteenth Century*. Auckland: Penguin Press, 1996.

Bellwood, Peter. *Man's Conquest of the Pacific*. New York: Oxford University Press, 1979.

Buck, Peter H. (Te Rangi Hiroa). *Vikings of the Sunrise*. Philadelphia: Lippincott, 1938.

Finney, Ben. *Sailing in the Wake of the Ancestors: Reviving Polynesian Voyaging*. Honolulu: Bishop Museum Press, 2003.

Fornander, Abraham. *An Account of the Polynesian Race, Its Origins and Migrations and the Ancient History of the Hawaiian People to the Times of Kamehameha I*, Vol. I of III. London: Trubner, 1878.

Heyerdahl, Thor. *Sea Routes to Polynesia*. London: George Allen and Unwin, 1968.

Horridge, Adrian. "The Austronesian Conquest of the Sea—Upwind." In *The Austronesians: Historical and Comparative Perspectives* (pp. 143–60). Edited by Peter Bellwood, James F. Fox, and Darrell Tryon. Canberra: ANU E Press, 2006.

Howe, K. R. *The Quest for Origins: Who First Discovered and Settled New Zealand and the Pacific Islands?* Auckland: Penguin, 2003.

Langdon, Robert. *Kon-Tiki Revisited*. Melbourne: Australian Scholarly Publishing, 2009.

Sharp, Andrew. *Ancient Voyagers in the Pacific*. Wellington: Polynesian Society, 1956.

Simmons, D. R. "A New Zealand Myth. Kupe, Toi and the 'Fleet'." *New Zealand Journal of History* 3, 1 (1969): 14–31.

Simmons, D. R. *The Great New Zealand Myth: A Study of the Discovery and Origin Traditions of the Maori*. Wellington: Reed, 1976.

Smith, S. Percy. *Hawaiki: The Original Home of the Maori; with a Sketch of Polynesian History*. Auckland: Whitcombe and Tombs, 1910.

13

The Toltecs and Maya developed wheels for religious reasons, but not for wheelbarrows or other practical uses. The reason is that they had sufficient slave labor.

PRO   Talaat Shehata
CON   Harald Haarmann

PRO

Although the Toltecs and Mayans made use of the wheel in calendars and other religious symbols, they did not use the wheel for practical purposes, in the ways that the earliest Mesopotamian, Egyptian, Sumerian, and other Middle Eastern societies did. However, there were other widespread technologies for which they had no use. Archaeological evidence has proven that the Toltecs did not have a practical use for writing alphabetically, and neither the Toltecs nor the Mayans used any forms of metal tools until 800 CE, money for bartering purposes, or the use of ancient and well-recognized and regarded beasts of burden, as donkeys, mules, oxen, camels, or horses. They simply had a historically parallel and unique way of going about developing their civilizations. The Toltecs and Mayans have historically appeared to stand on their own separate and unique grounds. The facts inform us that the Toltecs and Mayans brought forth, separately, *authentic* and very dynamic and prosperous civilizations in what is regarded as Mesoamerica, or what are current central Mexico, Guatemala, and Honduras.

Wooden, stone, and obsidian tools were used, along with extensive human labor, to build the magnificent architectural and temple structures that have been left behind. The wheel, the central harbinger of other Mediterranean, Asian, and Old World European civilizations, was only to be found in Toltec and Mayan relics as symbolic religious and cosmological items, or as toy objects to be enjoyed by their young or young at heart, especially during their joyous days of festivities. Besides being made of clay and containing miniature clay-rollers; the wheels of the Toltecs and Mayans were not later created for any constructive technological purpose. The slave argument—that by having slaves, the Toltecs and Mayans did not need to be primarily focused on developing the dynamic, mobile, and progressive nature of the wheel—seems to not fare well with the fact that the other, mostly Middle Eastern civilizations that most heavily used the wheel also made extensive use of slaves. Therefore, historically, what we need to

281

Mayan wheeled animal toy, found near Vera Cruz, Mexico. (Private Collection/Boltin Picture Library/The Bridgeman Art Library)

focus on is not so much why the Toltecs and Mayans did not make use of the wheel, as other global civilizations had, but, instead, how they were able to create the magnificent civilizations they did without the use of the wheel and other known Old World metal tools. Finally, what does it tell us about how civilizations had evolved in parallel paths throughout history, and continue to do so? Was and is there a predetermined manner and pattern by which they needed to adapt themselves, or was it ultimately of any historical significance that they could achieve the magnificent scales they had attained in their precise "moment" in history?

From the 10th to the 12th centuries CE, the Toltecs had achieved relative dominance over their neighbors in the central highlands of Mexico. Their capital was located in Tula. The legend of Quetzalcoatl, an ancient prophet ruler and divine presence to be emulated by future generations of Toltecs, Mayans, and Aztecs who would later conquer and subjugate the Toltecs for a few long centuries *prior* to the arrival of the Spanish and Hernando Cortez, played an important role in these societies' collective religious beliefs, traditions, and practices. A divine personification of absolute love, Quetzalcoatl, or the Feathered Serpent as he was often called by his devoted followers, was later misrepresented by the Aztecs in their daily and ritual practices, with the wholesale slaughtering and sacrificing of captive elite and a few nonelite prisoners and members of their own expansive civilization. This perversion of the original message of Quetzalcoatl was similar in scope to the Spanish perversion of Christ's message of love; they put many of the Native American inhabitants they encountered in the Aztec kingdom and beyond to the sword, in the name of Christ and his followers' ultimate "salvation."

Regarded as a "white bearded man" (which explains why the Spanish did not experience *initial* resistance to their presence when they first made contact with the Aztecs and the Native American inhabitants), Quetzalcoatl, a divine man of love and wisdom, issued strict rules against blood sacrificial practices, which had been rampant among the Toltecs and other tribal groups within the central Mexican highlands for centuries *prior* to his arrival. He exalted the presence of a single all-encompassing supreme being and introduced the use of the calendar to the Toltecs to better help them maximize the annual yields from

their staple crops, such as corn, squash, gourds, potatoes, and lima beans, to mention a few. A solar calendar of 365 days was put to use, and a lesser efficient ritual standard calendar of 260 days was also used.

From 1122 to 1150, the Toltecs were of the strong conviction that the reincarnation of Quetzalcoatl in the being of the son of Ce Tecpatl Mixcoatl, who had ruled Culhuacan during that period, had been realized in Ce Acatl Topiltzin. Topiltzin, within a few short years of his father's passing, gathered a small force and conquered the Toltec capital in Tula and publicly claimed that divine and royal title. During his reign, prosperity seemed to flourish throughout the land. The arts, extensive small metal industries, and diverse crafts took root and thrived in the larger society of around 120,000 inhabitants. Topiltzin, as the reincarnation of the divine presence of Quetzalcoatl, ceased the ritual practice of animal and human sacrifice. Most, if not all, domestic and neighboring violence by Toltecs was temporarily ended. But this unusual period of stability and peace would not last for long. By 1168, the Toltec civilization had reached its weakest point and gradually began to collapse. Folklore and the reputed *Annals of Cuauhtitlan* refer to sorcery being practiced on Quetzalcoatl, with the initial intent of getting him to change his policies regarding animal and human sacrifices. This included the uberdrive format by which a few imaginative members of the community, as in most other global cultures, identified Quetzalcoatl troubles with the "evil" deity Tezcatlipoca. As the story goes, since Quetzalcoatl refused to accede and return to the old bloody sacrificial practices of the magicians and priests that surrounded him, they were able to ally themselves with the powers of the "evil" deity Tezcatlipoca and have Quetzalcoatl escape Tula after being humiliated by Tezcatlipoca. How that humiliation was precisely undertaken is left conveniently "mysterious" in the *Annals* and the rendered folktales, which gives them the quality of gossip. To add more relish to the prefabricated tale, Quetzalcoatl, then in exile, decided to set himself on fire so that he would reemerge as Venus, the morning star. But, sadly, these elaborate tales never really established who the outside invaders were (though over time archaeologists, historians, and other scholars believe that the Toltecs were eventually conquered by the rising Mayan civilization) who actually contributed to the Toltec collapse. At that point in Toltec history, matters and events take on an ethereal appearance for any serious scholar or reader. It is for that precise reason that one can better understand how difficult it can be, unlike in researching and reflecting on Mediterranean historical issues, for one to gain a deeper understanding of the important realities and difficulties that scholars of Atlantic history have contended with. That is, while pursuing and researching the earliest stages of Atlantic history, as demonstrated in this latest Toltec example of how their civilization eventually collapsed, one needs to be able to gain a much deeper understanding of all the important individual pieces of the Mesoamerican and ultimately the earliest period of the Atlantic world.

The question that often imposes itself on scholars within the region concerns the primary interaction and convergence of the tribal groups, kingdoms, and civilizations within the central Mexican highlands, at the height of the Toltec, Mayan, and Aztec periods. Hundreds of languages abounded, and exact population sizes and the precise locations of these population centers are still unknown variables. This, of course, includes the lack of true historical knowledge of the exact time period that jurisdiction had been exercised by one people upon another, and what the ultimate definitive effect was to both the conquered and the conquerors. The interregnum between the collapse of the Toltec civilization and its eventual conquest by the Mayans, and the spotty, if not downright blotched, history as to what immediately followed, clear indications of such a very serious dilemma.

Despite their eventual collapse, Toltec religious beliefs, values, rituals, and traditions would conquer all future civilizations in the central Mexican highland region. As mentioned, the worship of Quetzalcoatl and the adoption and implementation of his precepts and teachings were of paramount significance. Similar to the taming nature of Islam on Genghis Khan's children, grandchildren, and great-grandchildren, and their Mongol hordes on the steppes of Southwest Asia, which contributed to the emergence of the artistic wonders and brilliance of the Mogul Empire, Toltec religious beliefs profoundly impacted the conquering Mayan forces and their daily, communal, spiritual, and artistic evolution. Being much more prone to very militaristic and aggressive violent tendencies, the Mayans gradually began to implement and place their unique stamp on their conquered subjects' art forms. The tendency in the beginning was to view more warrior elite-dictated art forms of military figures and the glorification of their achievements for the larger Mayan community. In time, militaristic propaganda began to give way to much more complex and stylistic art forms, intricate hieroglyphic texts, and the creation of more innate art works that glorified the beauty of this world which the deceased might take with them into the afterlife. This new trend for the Mayans only further emphasized the natural and newly acquired intellectual attributes of their larger culture, instead of the often prefabricated and self-aggrandizing achievements of the elite warrior class. Eventually, the Mayans, instead of maintaining their earlier militaristic autocratic ways, evolved into a much more peaceful theocracy. This new worldview would change in the early 1400s, once they encountered the conquering Aztec forces.

At the peak of their civilization, Mayans owed much of their adapted agricultural practices, growth of their cities, and the acquisition of writing tools to most of the surrounding areas of the geographic locales they had conquered. Squash, beans, and corn were raised in the low south and southwest coastland areas. They had been grown in those areas since 3000 BCE. The creation of pottery and wheel-shaped items among the pottery was attained by 2500 BCE. Between 1500 and 1200 BCE, in time, villages and cities emerged in those lowland areas. Between 800 and 600 BCE, writing emerged among the Zapotecs in the Oaxaca

region. The first states evolved by 300 BCE. The Mayans were left with the simple task to either internalize, develop, envision, adapt, or customize the contributions that neighboring and surrounding groups had to offer them, or reject them and continue their earlier rigid militaristic mindset. They wisely chose the former. Archaeological evidence shows that *within* the Mayan areas proper, indications of the mastery of pottery and the emergence of villages began around 1000 BCE. Imposing architectural structures and designs began to take shape around 500 BCE. Finally, written texts began to appear in Mayan culture, around 400 BCE. With 15,000 inscriptions written solely on pottery and stone, and the sole mention of the role, deeds, and conquests of members of the royal and noble classes, very little knowledge is available to cast light on the daily lives and function of the average Mayan man, woman, and child.

The Mayan king donned the dual political role of head of state and the religious one of highest priest. He often presided over calendar and astronomical events, at which wheels were often used and exhibited as an important supplementary item, which established in its intricate design the cyclical evolution of the cosmos and in some cases the location of the morning star, Venus. By being capable of exhibiting to the Mayan nobility and population his mastery of the cosmic order, the Mayan king was able to convince the people that he was capable of guaranteeing their collective prosperity and well-being by his predictions of when and how much rain would fall to inundate the often potentially drought-stricken fields. In Mayan culture, therefore, the king was the critical medium through which the people could channel their hopes between themselves and the gods. In reality, as in any game of chance, the odds were greatly stacked against the player; in this case, the king. So, he spent most of his time thinking on his feet and guaranteeing that his nobles and military elites were loyal, consolidated, and forever showered with material and land ownership favors.

The large mass of the Mayan peasantry and population were only too happy to keep their king, his family, the advisers, and members of the nobility contentedly living in luxury. They not only built majestic architectural and often beautiful interior designed structures, palaces, and mansions for them, but also kept them well fed on venison meat, corn, squash, lima beans, and other assorted of beans and food items. But, curse the days that the king was unable to provide them with the much-needed rain for their crops and continue to maintain, if not increase, the prosperity that they felt was annually due them. This pattern of interaction between the king and his subjects, along with some very serious climatic change, extreme drought conditions, soil erosion, the significant increase in unusable fallow fields, and the profound increase in population as well as with some self-defeating ritualistic practices and outside foreign threats, led to instability. The Mayan civilization experienced a precipitous rise and fall, between the first empire, which lasted from 200 to 850 CE, and the second and last empire, which lasted from 1000 to 1350 CE. All this occurred with the internecine problems that

A Spanish Official Encounters the Mayan City of Palenque

Among the first Europeans to visit the magnificent Mayan city of Palenque was Antonio del Rio, who led an expedition there in 1784. Below is his description of the architecture he found there.

The interior of the large building is in a style of architecture strongly resembling the gothic, and from its rude and massive construction promises great durability. The entrance is on the eastern side, by a portico or corridor thirty-six varas or yards in length and three in breadth, supported by plain rectangular pillars, without either bases or pedestals, upon which there are square smooth stones of more than a foot in thickness forming an architrave, while on the exterior superficies are species of stucco shields, the designs of some of them, accompanying this report, are numbered 1, 2, 3, while, over these stones, there is another plain rectangular block, five feet long and six broad, extending over two of the pillars. Medallions or compartments in stucco containing different devices of the same material, appear as decorations to the chambers, . . . and it is presumable, from the vestiges of the heads which can still be traced, that they were the busts of a series of kings or lords to whom the natives were subject. Between the medallions there is a range of windows like niches, passing from one end of the wall to the other, some of them are square, some in the form of a Greek cross and others, which complete the cross, are square, being about two feet high and eight inches deep. . . . Beyond this corridor there is a square court, entered by a flight of seven steps; the north side is entirely in ruins, but sufficient traces remain to show that it once had a chamber and corridor similar to those on the eastern side, and which, continued entirely along the several angles. The south side has four small chambers with no other ornament than one or two little windows, like those already described. The western side is correspondent to its opposite in all respects, but in the variety of expression of the figures in stucco: these are much more rude and ridiculous than the others, and can only be attributed to the most uncultivated Indian capacity.—The device is a sort of grotesque mask with a crown and long beard like that of a goat, under this are two Greek crosses. . . .

It is by no means improbable that these fantastic forms, and others equally whimsical, were the delineations of some of their deities to whom they paid an idolatrous worship, consistent with their false belief and barbarous customs.

Source: Antonio del Rio. *Description of the Ruins of an Ancient City, Discovered Near Palenque, in the Kingdom of Guatemala, in Spanish America.* London: Henry Berthoud, 1822.

continued to plague them for over a millennium, which finally caught up with them and directly contributed to their collapse and defeat by the Aztecs, in the mid- to late 14th century CE.

Instead of simply viewing Mayan geographic conditions as being tropical or a rainforest, we need to understand that since much of their livelihood was only within a thousand miles of the equator, with 17 to 22 degree latitude readings

during the first four months of the year (January to April), their homeland experienced regular very dry months. As for the rainy season, that usually arrived from May through October of each year, which would be the only time a neutral observer might accurately regard it as a "seasonal tropical forest." So, the reality of the Toltec and then Mayan civilizations was that they had to adapt themselves to the extremes of their environments. Often, one slight annual misstep would lead to disaster for their people. This was never made any clearer than when the first few early and larger classic Mayan civilization collapses occurred by the ninth century and continued into the tenth century CE. During that time period, the steep hills were used for planting, which in comparison with the earlier used valley soil, were much more acidic in composition, less fertile, and had a much lower yield in phosphate retention and production.

It is a well-known fact among farmers in the central Mexican highlands that corn and other vegetable yields in the fertile valley fields are much higher and richer in protein and fiber content than anything that could be grown on the hill slopes. This was the ultimate dilemma that the Mayans had to contend with during their civilization's growth over the millennia. This fact amazes many archaeologists, historians, and scholars as to the Toltecs' and Mayans' individual and collective abilities to create, innovate, and continue to maintain their civilizations under such dire conditions for such a prolonged period of time without making use of the wheel or any of its later contributing technologies as parallel civilizations in the Middle East, Asia, and the Old European World had done.

During each of the separate collapses throughout their protracted history, the Mayans built larger and more intricate structures, temples, and palaces. The idea that preoccupied them was that when times were hard, you just tried to build bigger and better edifices to please the gods. It is estimated that from 250 CE, when their dire water, irrigation, drought, and climate change conditions were evident, the Mayans increased the construction of their monuments, city structures, and temple sites with their increased amounts of wheel-shaped pottery for cosmological and ritualistic uses, in exponential proportions. The greatest

Disc with relief design representing a ball player. Around the edge of the disc are a series of dates including day and 20-day period signs. Mayan, 590 CE. (Museo Nacional de Antropologia, Mexico City, Mexico/ Giraudon/The Bridgeman Art Library)

amount of construction often occurred a few short years prior to the actual fall of either of the civilizations in question, (i.e., the first major fall in the 9th and 10th centuries, and during the second and final one in the mid to late 14th century). The Mayan population had grown exponentially during both periods. The demand on the shrinking agricultural land had grown vicious, especially on the relatively infertile steep hill soil, which seemed to be the only land left for Mayans to compete for. With increased construction projects, sediment erosion of the hill and valley fields increased. That only added more to the Mayans' troubles. The hill slopes had become so eroded that whatever nutrients they still had available for future farming were being eradicated. With the rainy season, much of these sediments, combined with the initial acidic nature of the soil in the hills, were washed down into the more fertile valleys, creating its own chaos. By 700 CE the soil in both the valley and the hill slopes had become relatively toxic, too toxic for any future agricultural benefits by the Mayan population. Added to the mix was the regular use of plaster for their building projects and as writing objects, which the Mayans stripped off the barks of the surrounding trees in the forest. This, in time, created a looming deforestation catastrophe. Pine trees that had been growing in the central Mexican highlands for eons within a few short centuries had been completely cleared from all the hill slopes. The trees were not only used for construction and plastering purposes, but also for their fueling needs. This in time contributed to the increased drought cycle, since the presence of trees had helped continue whatever increase they enjoyed in extra rainwater. Now, with the forests gone, less rain fell to help irrigate their field on the hill slopes and in the valleys.

From 760 to 910 CE the Mayan collapse spread throughout its different power centers throughout the region, incrementally and at different stages. But, the collapse of the civilization that the Mayans had known for over a millennium had arrived and was merciless. The results were so glaring that the overall impact on such a flowering and productive civilization was sobering to any serious reader or scholar's eyes. From what had been the most productive, overpopulated, artistic, and most vibrant regions of the Mayan civilization, the low southland area, during the period of the first Mayan civilization's fall, 99 percent of its inhabitants had disappeared. Most perished as a result of starvation, thirst, and regular killings of one another because of conflicts over continued limited resources. With increased warfare among themselves, the scarce resources growing scarcer, and shrinking land spaces surrounding them, the extremely valuable properties still available evolved into no-war parches of land that the different communities were forced to honor in the attempt to maintain a zone of insulation from one another. This continued the downward spiral of no further agricultural land to farm. Yet another important factor to keep in mind is that with the exponential increase in population and no more land to occupy once the old agricultural property had turned fallow or become drought-stricken, the Mayans found themselves with their individual and collective backs very seriously up the irascible wall. Each Mayan was forced into the unenviable position of making a last

stand in order to survive. What's even more sobering was that they were faced with those deadly nonchoices, once again, by the time of the second major Mayan collapse, when the Spanish had made first contact with them in the first decade of their arrival in the early 1500s. A population of over 33 million Mayans had shrunk to a mere 30,000. Which leaves us with this very serious question: Could current postmodern and modernizing societies and future generations learn from this hard lesson and not allow it to happen to them? History will definitely keep us informed.

Besides all the unique and astounding features of both Toltec and Mayan civilizations, the important lesson that should be drawn from their existence and continued persistence over the centuries, against often insurmountable geographic, physical, and ecological odds, is not that they never "constructively" made use of the wheel or what different technologies it was able to spawn over the centuries in the early Middle East, Asian, and Old World societies; but, instead, could any of these wheel-driven societies have dealt with the internecine geographic, physical, and ecological problems that the Toltecs and Mayans daily dealt with and arrived at any better results? It's very doubtful. The wheel question, despite its importance in historical context when one looks at the larger landscape, seems quite trivial. In current postmodern advanced societies, it's sad to see that we are dealing with the same problems, if not much more serious ones, in matters as climate change, green-house gas emissions and their catastrophic impact on the Arctic region, the erosion and depletion of prime agricultural soil and land, a persistent exponential increase in the global population, dwindling water resources, an increase in the civil wars on a global basis over limited natural resources, and growing rates of domestic crimes, abuse, violence, and warfare throughout the world. The wheel, in perspective, hasn't really done anything to address, if not alleviate, any of these very serious problems. They continue with us today, as they did during the height of the Toltec and Mayan civilizations; except, today, and into the near and distant future, it has become much more magnified. If there's anything to be learned from the Toltec and Mayan civilizations, besides how were they able to create and maintain their separate civilizations for numerous centuries, without the use of Middle East, Asian, or Old World technologies, it is that we must avert the manmade and nature-created catastrophes that they dealt with before we find ourselves meeting the same ultimate fate that they met, after an extended flowered and magnificent period of existence.

References and Further Reading

Bailyn, Bernard. *Atlantic History: Concept and Contours*. Cambridge, MA: Harvard University Press, 2005.

Bentley, Jerry H. "Cross-Cultural Interaction and Periodization in World History." *AHR* 101 (June 1996): 749–70.

Demerest, Arthur, Prudence Rice, and Don Rice, eds. *The Terminal Classic in the Maya Lowlands*. Boulder: University Press of Colorado, 2004.

Fagan, Brian. *The Long Summer: How Climate Changed Civilization*. New York: Basic Books, 2004.

Gill, Richard. *The Great Maya Droughts*. Albuquerque: University of New Mexico Press, 2000.

Lentz, David, ed. *Imperfect Balance: Landscape Transformation in the Pre-Columbian Americas*. New York: Columbia University Press, 2000.

Lerner, Daniel. *The Passing of Traditional Society*. Glencoe, IL: Free Press, 1958.

Manning, Patrick. "The Problem of Interactions in World History." *American Historical Review* 101 (June 1996): 777–81.

Pilcher, Jeffrey M. *Food in World History*. New York: Routledge, 2006.

Redman, Charles. *Human Impact on Ancient Environments*. Tucson: University of Arizona Press, 1999.

Sharer, Robert. *The Ancient Maya*. Stanford, CA: Stanford University Press, 1994.

Turner II, B. L. "Prehistoric Intensive Agriculture in the Mayan Lowlands." S*cience* 185 (1974): 118–24.

Organski, A. F. K. *The Stages of Political Development*. New York: Knopf, 1965.

Webster, David. *The Fall of the Ancient Maya*. New York: Thames and Hudson, 2002.

Whitmore, Thomas, and B. L. Turner II. "Landscapes of Cultivation in Mesoamerica on the Eve of the Conquest." *Annals of the Association of American Geographers* 82 (1992): 402–25.

Worster, Donald. "World without Borders: The Internationalizing of Environmental History." In *Environmental History: Critical Issues in Comparative Perspective*. Edited by Kendall E. Bailes. Lanham, MD: University Press of America, 1984.

CON

The web of ideas that are placed in opposition on whether the Toltecs and Mayan developed the wheel for religious reasons versus practical uses related to the presence or absence of beasts of burden or slave labor is highly arbitrary, and its logical foundation is disputable. What does the practical use of the wheel have to do with slave labor? Is it reasonable to explain the absence of the one (i.e., the wheel to facilitate labor) with the abundance of the other (i.e., slave

labor)? No, it is not and there is good reason to reject this opposition as pseudo-logical. In this context, the existence of the one factor cannot explain the absence of the other. This lack of explanatory potential can be illustrated by prominent instances of cultural history.

For example, to those who, in the early civilizations of the Old World, carried out the monumental building projects—of the ziggurats, the stepped temples in Mesopotamia, and of the pyramids of the Old Egyptian Kingdom—the wheel was known as a practical device, and it was used. The invention of the wheel in the Old World was an independent event, and this event was not related to the availability of a human workforce. The wheel had already been in practical use hundreds of years before the first monuments were erected.

The monumental structures in Egypt and Mesopotamia could have been erected with the exclusive help of manpower. The corresponding large structures (i.e., ceremonial platforms, pyramids, temples) of the pre-Columbian era in the New World demonstrate that the construction of large-scale architecture without the use of the wheel was possible. The fact that the workforce in the Old World civilizations that were communal or slave labor-oriented had the wheel at their disposal made work much more effective, but the completion of the projects, in theory and practice, did not require the interplay of the two factors (i.e., manpower + wheel) as a necessary precondition.

The Mesoamerican "Missing Link"

In the New World, too, the wheel could have been introduced as a practical device at any time. At least the preconditions for such a move were present. Inventive thinking had been widely applied for various technical skills (e.g., masonry, pottery-making, mining, the working of hard stone such as jade, and rubber production). The idea of a turning wheel or disk was known as a symbol from the American calendrical system. The historical stage was set for a combination of the idea of the wheel and the application as a device by technical skills (an option that did not materialize).

Is there something like a "missing link" between the idea of the wheel as a cognitive concept and its practical application as a device for transport, a clue that existed in the Old World but was missing in the New World? In terms of technical skills, the pre-Columbian Mesoamericans could compete with any of the other ancient peoples of the world. They even invented techniques to work the hardest stones that exist.

> Jade was the most highly prized mineral in Pre-Columbian Mesoamerica. . . . Jadeite is much harder than obsidian and slightly softer than quartz; because of its structure, it is the toughest and most durable of stones. . . . The ability to carve and polish extremely hard stones seems to have been known before

the introduction of the Olmec style . . . but the Gulf Coast Olmec may have perfected the skill. (Pohorilenko 1996: 120)

The synchronization of the invention of wheel and the mobilization of a large workforce would not need to be seen as working the same way in America as it did in the Old World. World history teaches valuable lessons, provided one opens one's mind to learn them. The lesson of the wheel is highly instructive in that it illustrates the fundamental insight that there is no automatism in cultural evolution. The processes of technical innovation that operated in the Old World did not operate in the same way in the New World.

A special challenge must have been felt when, in the second millennium BCE, a network of long-distance transport without wheelbarrows was established (see below). Nevertheless, the idea of the wheel was never explored in Mesoamerica for possibilities of its practical manifestation. There was a "missing link" in the chain from idea to application, and because there was something missing, this caused a blocking of practical application.

The wheel is not the only issue that may illustrate fundamental differences of development. Another issue is the absence of iron melting in pre-Columbian Mesoamerica. It may always remain a mystery why the Native Americans mined iron ore but did not process it by applying melting techniques to extract the metal. Instead, they worked the ore (i.e., magnetite) and molded and polished it to produce mirrors. There are reasons why the development of the wheel differed in America from Asia, Europe, and Africa. The question why the pre-Columbian Americans did not use the wheel for practical purposes is usually discussed in relation to the aspect of the use of the wheel, that is from a standpoint where "practical use" is placed in opposition to "nonpractical use." This approach is biased, because a mode of thinking that gives priority to considerations of utility and mundane functions of technical innovations is typical of our time, and of the Euro-American worldview in particular. Such thinking was definitely absent from life in the Mesoamerican communities of the pre-Columbian era.

To understand the ways of the pre-Columbian Americans and their worldview, we are advised to refrain from any projection of our modern views onto their world. What is called for is an internal reconstruction of pre-Columbian realities of community life, whereby paying due tribute to the conceptualizations of the American ancestors, as can be concluded from their cultural heritage. This means that the modern investigator has to make an effort to "stand beside him- or herself" and to become familiar with the mindset of those who created the pre-Columbian civilizations to avoid distortion by modern ideologies. Any discourse about the significance of the wheel in those remote cultures is only meaningful within the context of a reconstructed worldview.

Cultural Trajectories of the Mesoamerican Mindset in the Pre-Columbian Era

The discussion of differences of mindset may start with the most unifying of all cultural concepts—how human beings in different cultures perceive the idea of "community." According to the pre-Columbian mindset, the venerated ancestors as well as the representatives of the living generations were members of the community. The Maya buried their dead beneath their houses as an expression of a close relationship between the venerated dead and the living.

> The link between the community, Earth, and the ancestors was manifested in other ways too. The dead could be shown as trees planted in the soil, as on the sides of the sarcophagus in the Temple of the Inscriptions at Palenque. Here the ancestors of the ruler Pacal are shown as trees emerging from fissures in the earth. (Marcus 2000: 237).

Venerated ancestors were imagined as trees that were important for the Maya diet and economy. Cacao, chicozapote, avocado, guayaba, coyol, and mamey trees have all been identified as species of this category of "ancestor tree." As the reference to the tree spirits of ancestors illustrates, the Maya imagined their community as an extension of the space for the living, which also included spirited nature.

Since the times of the Olmec civilization, which had already emerged in the second millennium BCE, daily life of the pre-Columbian Mesoamericans unfolded under the auspices of a balance between the world of humans and that of the spirits. The cosmos of the inhabitants of the agrarian village communities comprised its own cultural living space; the surrounding nature, which was imagined to be spirited; the ancestors whose spirits were believed to advise and guide their living descendants in their mundane matters; and the supernatural beings (i.e., divinities and their various functions).

> For the indigenous person [i.e., the pre-Columbian Mesoamerican], the natural world was not something to be manipulated, exploited, and destroyed by humans at their pleasure, as it has been for Western culture. Rather it was a sphere populated by supernatural powers and forces with which people had to forge ties, necessary for the survival of humankind and for the conservation of nature, and as the context in which sacred beings manifested themselves. (de la Garza 2000: 70)

The balance between the worlds had to be maintained through rituals and ceremonies. The ceremonial traditions of the Mesoamericans are manifold, and many have persisted, in various transformations, over many hundreds of years to the

present. A colorful example of the perpetuation of ancient beliefs is the tradition to celebrate the "Days of the Dead" in Mexico on the occasion of All Saints' Day at the beginning of November. People would visit the graveyards, decorate the graves, and leave food and drinks for the ancestors, and then they would invite the ancestors' spirits to their homes and organize a feast for them. After the celebration, they would escort the spirits back to the graveyard to their resting place.

The mindset of those people who lived in the spirited world of pre-Columbian cultures lacked something that is so typical of the Euro-American way of thinking: striving for innovations and inventions for the sake of progress in technology. When reference is made here to the absence of a quality of the mindset, this does not mean that the pre-Columbian Mesoamericans lacked inventiveness. On the contrary, those who crafted the ancient American civilizations possessed astounding skills and mental capacities as well as technical know-how. Their skills were embedded in a worldview in which all technical progress was measured in terms of its benefit for the ritual balance between humans and the spirit world. One aspect that was definitely negligible in the thinking of Mesoamericans was the idea to save energy and manpower through the introduction of technical innovations to make daily work more efficient.

A way of thinking without the monopoly of practical priorities did not support technical inventions of the kind that were made in the Old World, where a different mindset might have dominated. This mindset must have been more problem-oriented toward devising practical solutions for things and toward improving existing technologies. Such thinking was the mother of many inventions in the Old World, among them the use of the wheel for work and transport. The first wheeled wagons appear on the western periphery of the Russian steppe zone about the middle of the fifth millennium BCE. About that time, the potter's wheel was introduced in Europe (Ukraine) and in western Asia (Mesopotamia).

Cultural Relativity as to How the Abstract Mind Works

We modern people identify objects according to their form and shape, not necessarily according to their function. In this way, we perceive what is similar or identical from the outer appearance. An object that is round and has an axis hole in the middle is called a wheel. In English, the same term wheel is used, regardless whether we speak of a wheel on a car or as part of a motorcycle, of a potter's wheel, of a wheel-shaped part in any kind of machinery, or of the picture of a wheel. Since the same term is used for the most different contexts in which a wheel-shaped object may appear, our abstract mind easily perceives all these objects as similar or identical. When we look at whatever wheel-shaped object, we readily associate the practical use of the wheel in our daily life, and we do this because our minds are conditioned by knowledge about the practical functions of the wheel that has been transferred over many generations in our cultural memory.

Only if we inspect the relationship of wheel-shaped objects and the terminology that is associated with them in other languages will we notice differences in the perception of such objects in the minds of people who live in cultural environments distinct from the world of English. One such environment is Tibetan culture. The Tibetans adopted Buddhism in the early Middle Ages, and, together with the scriptures and the teachings, they became acquainted with the eight-spoked wheel, the sacred symbol of infinite spiritual power and energy. The term for the symbol of infinite spiritual energy is *chakra*, a word borrowed from Sanskrit that was introduced to Tibet as an element of Buddhist worldview. When Tibetans are asked how they perceive the meaning of *chakra*, their answers indicate that this religious symbol is stored in the minds of Buddhists in a "separate chamber" and that all possible connotations range in the domain of abstractness. This means that it is difficult for a Tibetan to think of this same wheel as being a practical device for transport. The conceptual difference between the religious wheel and the wheel for practical functions is anchored in the language. The expression for a wheel as a device for transport is completely different: *korlo*. In the marked difference of items of the linguistic matrix is reflected an equally marked distinction of concepts on the cognitive level (i.e., wheel 1 = religious symbol vs. wheel 2 = practical device). Such contrasts like those described for Tibetan culture and language also exist in other cultural environments. For instance, the term for a potter's wheel is *durgn* in Armenian. Since this expression deviates clearly from other words for wheel-shaped objects, it can be concluded that Armenians, guided by their language in their perception, do not necessarily identify all wheel-shaped objects as similar or identical.

For the pre-Columbian cultural horizon, such distinctions cannot be demonstrated for the simple reason that the wheel for practical functions was never introduced, and the only concept that was associated with a wheel-shaped object was the wheel in a religious context. In the ancient cultures of the Old World, the wheel as an abstract motif is known from imagery dating to periods before the invention and the introduction of the device for transport. Therefore, in the Old World, the conceptual spectrum for wheel-shaped objects was extended to its full range. In the New World, this range was not explored in its total extension.

The Wheel as a Religious Symbol in Pre-Columbian Mesoamerica

Like people all over the world, the Native Americans knew what wheel-shaped objects looked like because of their observations of natural phenomena. The most impressive wheel-shaped objects in the sky are the sun and the moon. The latter celestial body attracted special attention because of its shape-shifting rhythm: full moon, descending crescent after 7 days; ascending crescent after 14 days; full moon after 7 days. It is noteworthy that the motif of the disk (or wheel or circle) is among the oldest ornaments of Olmec art, and it had already

featured on sculptures dating to ca. 1150 BCE. In numerous transformations, this basic concept persisted throughout the periods of pre-Columbian cultural development, and the disk/circle/wheel motif is, in manifold ways, combined with other motifs in Mesoamerican religious iconography. In the Mayan tradition, the icon for the sun was a four-petaled motif, which is known as the *kin* sign. *Kin* in Mayan means both "sun" and "day."

The sun and the moon were conceived as divinities, like other known celestial bodies. However, these two bodies, which impressed the Mesoamerican mind by their size and properties, held a prominent position among all the other gods. The sun was personified as a male god, perhaps because of the vigor and strength of the solar energy, especially when thinking of the power of the rising sun. Contrasting with this personification, pre-Columbian mythology and art identified the moon as feminine. The duality of the sun god and the moon goddess manifests itself in many monumental buildings dedicated to these divinities. Perhaps the most impressive ensemble is that which constitutes the two pyramids, the larger sun pyramid and the smaller moon pyramid, in the ancient town of Teotihuacan, which was called "city of the gods" by the Aztecs. The civilization of Teotihuacan, situated some 40 km to the northeast of Mexico City, flourished between ca. 100 and 600 CE. The city was abandoned, for unknown reasons, in the early ninth century, but it was later frequented as a center of pilgrimage by the Aztecs.

The name of the sun god in Mayan is *Kinich Ahau* ("sun-faced or sun-eyed lord"). In later Mayan iconography, "the sun god is closely identified with jaguars, and at times appears with a jaguar ear" (Miller and Taube 1993: 106). In the Aztec religion, the sun was personified as *Huitzilopochtli*. For the Aztecs, this male figure played the role of a supreme god, and his attributes, which he holds in his hands, are a petaled disk and the fire serpent. The Mayan moon goddess is often identified as *Ixchel*, a figure whose name can be translated as "Lady Rainbow." However, the image of the moon goddess shows a young and beautiful woman, while *Ixchel* was depicted as an old woman. The identification of the moon goddess with *Ixchel* may have resulted from confusion caused by the description of the Mayan gods in the early Spanish accounts of the native Mesoamericans in the sixteenth century. Then, *Ixchel* was the most prominent figure among the Mayan. The original name of the Mayan moon goddess is not known. She is usually depicted as sitting on the crescent, which is the glyph for moon in Mayan writing. Her typical attribute is a rabbit. According to classical Mayan beliefs, the picture of a rabbit becomes visible on the surface of the celestial body at the time of the full moon. The moon goddess was held in high esteem, and she was venerated as the patroness of pregnancy and childbirth, of weaving and divination, and also of fertility. In the Aztec mythical tradition, the moon goddess is called *Coyolxauhqui*, and she is identified as Huitzilopochtli's sister.

Extensive ceremonial services were carried out to worship both the sun and the moon with their prominent status in the Mesoamerican pantheon. In view of the sanctity of the two concepts of the sun and the moon and of the attention these two celestial bodies and their personifications enjoyed in the pre-Columbian communities, it is reasonable to assume that their visual images were also sanctified. Once the shape of the disk or circle is sanctified as an attribute of a prominent divinity, the cognitive path to imagine any practical use for a device in the same shape as a divine attribute is most likely to be blocked. This blocking, postulated here as a psychological phenomenon, can hardly be evidenced with any certainty for the mindset of people living in bygone cultures. And yet this kind of psychological impediment to explore the benefits of wheel-shaped devices cannot be reasonably ruled out. In this context, it is amazing that some of the pre-Columbian imagery illustrates wheel-shaped motifs that spontaneously evoke the association with spoked wheels on wheelbarrows or wagons in the mind of Westerners.

The Role of the Calendar Wheel

The impression of the changing of the seasons of the year must have spurred reflections about the vegetation cycle among the pre-Columbian Mesoamericans in the early agrarian communities of the second millennium BCE. For the Mesoamerican calendar system, though, other aspects of cyclic events obviously had greater importance than the vegetation cycle. This can be conjectured from the number of days that were counted in the oldest known Mesoamerican calendar system, the so-called 260-day almanac. The specific number of 260 days is not associated with either astronomical or agricultural phenomena. This system "was probably devised by midwives to calculate birthdates, working from first missed menstrual period to birth, approximating the 9-month human gestation period" (Miller and Taube 1993: 48).

The oldest evidence for the pre-Columbian calendar system, which originated in the first millennium BCE, comes from the Olmec civilization in the coastal areas of the Gulf of Mexico. The earliest calendrical inscription dates to the sixth century. This calendar persisted in religious functions throughout the pre-Columbian period. Its name in Mayan has been reconstructed as *tzolkin*, and the Aztecs called it *tonalpohualli*. A civic calendar was used alongside the ritual calendar for daily use. This calendar counted 365 days and was called *haab* by the classical Maya. According to the Mayan tradition, dates were always given in the form of a double count, thus synchronically counting an event based on both calendars and their time measurements. The operation of the calendar was conceptualized in terms of the turning of a wheel. Since there were two parallel calendrical systems, there were two calendar wheels that turned synchronically. It would take a long span of time—or exactly 52 years of 365 days (= 18,980 days)—so that a given date would repeat itself. This period is called the calendar round.

A Mayan Wheel in England

One of the most interesting and controversial appearances of a Mayan wheel took place centuries after the decline of Mayan civilization, and thousands of miles away. During August 2–3, 2004, a large crop circle in the shape and design of a Mayan wheel appeared at Silbury Hill, Wiltshire, England. The wheel, used by the Mayans as a calendar, counts down to the year 2012, and some "experts" have ominously predicted that this seemingly supernatural apparition portends the end of the world. Of course, the fact that 2012 is the end point of the calendar is no surprise to anyone who has studied actual Mayan calendars, as they all indicate an end-of-time point in the year 2012.

The wheel as a visual icon to illustrate the proceeding of the time count according to the two calendars became a common motif in pre-Columbian iconography and architecture. Well known is the Aztec calendar stone in the form of a big wheel, and the archaeological record yields numerous disk-shaped stones with signs of calendrical notation. The familiar motif of the calendar wheel was also reproduced in some of the late Mayan books and in early Spanish manuscripts of the 16th century in which the Native American calendar system was documented at a time when it was still in use. In a highly pervasive way, the visual impression of the calendar wheel illustrates the closeness of ideas, of the abstraction of the cyclical movement of time counting, and of the wheel as a practical device. Despite the familiarity and the popularity of the calendar wheel as an (abstractly) turning device, this association never spurred the pre-Columbian mind to transfer the idea, in a process of cognitive analogy, and to make the device work for mundane purposes for transport and labor.

Trade Routes and Transport

The pre-Columbian civilizations all share basic properties—of cultural patterns (e.g., the calendrical system), of ornamentation in the visual arts and architecture (e.g., the jaguar motif), of belief systems (e.g., the cult of the rain god), and of ideas about life (e.g., the cosmic cycles of a renewal of life). Scholars have long wondered how cultural traditions and technical skills could have spread so widely over a large area and what the origins were. Modern research has produced insights that allow for the reconstruction of a sophisticated network of trade relations throughout Mesoamerica and of a web of idea diffusion. The Olmecs were the first Native Americans who explored trade routes leading from the Gulf Coast inland. These routes, which were traveled from the second

millennium BCE, connected the coasts of the Atlantic with those of the Pacific. It is over this network of trade and communication that the foundations of the classical civilizations were established.

Manifold goods were traded, among them cacao (chocolate), mollusk shells, turtle shells (for making drums), fish, stingray spines, shark teeth, rubber, salt, tar, pottery, clay, obsidian, iron ore and pigments, turquoise, mica, and so forth. Among the luxury goods in the Olmec-controlled trade from coast to coast were spondylous shells, pearl oyster, jade, and alabaster. Trade was carried out between villages, but the routes that were regularly traveled also included connections with areas that were important for the exploitation of raw material such as obsidian (used for tools and ornaments), precious stones (e.g., jade, serpentine), and minerals (e.g., magnetite for making mirrors).

Manpower was the only resource throughout the pre-Columbian era to keep up trade relations and to guarantee the movement of materials and commodities. Beasts of burden were unknown as were wheelbarrows. Transporting raw material such as iron ore and stones (i.e., obsidian, jade), ceramic objects, shells, and foodstuff only by way of carrying everything on one's back must have been laborious. The lack of beasts of burden is no criterion to pervasively explain the lack of the wheel. Wheelbarrows can be drawn by humans, applying the principle of transport that is well known from Southeast Asia (i.e., transportation by means of rickshaw carts).

The light of adapting the idea of the wheel to make it work as a practical device obviously never flared up, because the incentive to surpass the impediment of a religious "blocking" was never activated. And yet, whatever approach may be chosen to explain the missing link, none of these approaches are ultimately satisfactory. So the absence of the wheel in practical functions may always remain a mystery of pre-Columbian history.

References and Further Reading

Adams, Richard E. W. *Ancient Civilizations of the New World*. Boulder, CO: Westview, 1997.

de la Garza, Mercedes. *The Mayas: 3000 years of Civilization*. Mexico City: Monclem Ediciones, 2000.

Diehl, Richard A. "Tula and the Tolteca." In *The Aztec Empire* (pp. 124–27). Edited by Felipe Solis. New York: Solomon R. Guggenheim Foundation, 2004.

Haarmann, Harald. *Foundations of Culture. Knowledge-construction, Belief Systems and Worldview in Their Dynamic Interplay*. New York: Peter Lang, 2007.

Marcus, Joyce. "Toward an Archaeology of Communities." In *The Archaeology of Communities. A New World Perspective* (pp. 231–42). Edited by Marcello A. Canuto and Jason Yaeger. New York: Routledge, 2000.

Miller, Mary, and Karl Taube. *The Gods and Symbols of Ancient Mexico and the Maya. An illustrated Dictionary of Mesoamerican Religion.* London: Thames and Hudson, 1993.

Pohorilenko, Anatole. "Portable Carvings in the Olmec Style." In *Olmec Art of Ancient Mexico* (pp. 119–31). Edited by Elizabeth P. Benson and Beatriz de la Fuente. Washington DC: National Gallery of Art; New York: Harry N. Abrams, 1996.

14

Native American languages can be traced to three grand linguistic roots.

PRO   Harald Haarmann

CON   Peter N. Jones

PRO

The history of surveys and classifications of Native American (Amerindian) languages reaches as far back as the 17th century. No other approach to a historical classification has spurred as lively a debate among linguists, anthropologists, and archaeologists as has the approach presented by Joseph H. Greenberg in 1987, who distinguished three macrophyla, or linguistic groups. First, the Amerind macrophylum, comprising the great majority of indigenous languages of the Americas, which are grouped in 11 subfamilies. The more than 900 individual languages of this macrophylum are assumed to be descendants of a common basis (protolanguage) that was transferred to America with the first wave of migrants from Siberia. The conventional date for that migration is given as some 13,000 years ago. Second, the Na-Dene languages, comprising the Athabascan language family (Navajo, Apache, etc.) and several language isolates (Eyak, Tlingit, Haida) in the northwestern part of North America. The common basis for the languages of this macrophylum (altogether 42 individual languages) was transferred to America by the migrants of the second wave who came some 11,000 years ago. Finally, the Eskimo-Aleut languages, comprising local variants of Eskimo (Yupik, Inuktitut, Inupiatun, Inuit, etc.) and of Aleut. These languages have derived from a common basis that is about 9,000 to 10,000 years old. The evidence presented here will prove that Greenberg's ideas on the basic macrophyla of Amerindian languages is correct.

This distinction of linguistic macrophyla is in accordance with the three major migrations in the course of the prehistoric peopling of the Americas that have been identified by archaeology and human genetics. While the historical relationships within the macrophylum of the Na-Dene languages (and of the Eskimo-Aleut macrophylum, respectively) are undisputed, most scholars in the field of linguistics oppose the higher-order classification of the Amerind stock and postulate a greater number of language families. In the ongoing controversy, marked positions of pro (represented by the "geneticists") and con (propagated by the "diffusionists") are taken.

301

The overall pan-American vision of the linguistic landscape as propagated by Greenberg has almost been buried under the critique of shortcomings of his methodology brought forward by the diffusionists. Despite a continuous dispute over the reliability of compared lexical items, reconstructed forms, and the classification of individual languages, the main issue has remained, in principle, one of perspective. If one follows the great currents of the early settlement of America and the early movements of populations that have been reconstructed by human genetics, then it seems conclusive to strive to reconcile findings of historical linguistics with these insights. That is Greenberg's perspective. What is reflected in the critique by diffusionists of Greenberg's comparisons and categorizations is the state of agony in which historical linguistics finds itself with its rather insufficient methodology to reach deep beneath the horizon of time, rather than the negation of sets of lexical equations that are historically related to certain postulated cognate words as their common basis. In the following outline, the emphasis is more on the relativity of perspective and methodology (linguistic versus nonlinguistic) and less on the discussion of details.

Anyone who engages in the debate about language classification in the Americas has to cope with natural limitations of the documentation of the subject matter, the some 1,010 native languages. Although the amount of data about the Amerindian languages is continuously growing, grammatical descriptions and dictionaries are, by far, not available for all languages. Also, for practical purposes any comparative study has to limit itself to a selection of analyzed languages and a selection of vocabulary. The some 2,000 words in Greenberg's catalog of cognates are but a fraction of the entire lexicon of any living language. Nevertheless, Greenberg's overview is the most comprehensive of all the lists that have ever been applied by comparative taxonomy. Greenberg has been criticized, notably by Ives Goddard (1987), for shortcomings in the reconstruction of historical protoforms of the Amerind languages. Here, the critique seems to be at odds with the possibilities to explore deeper chronological layers in the evolution of languages in convincing ways with comparative-analytical methods. In this context, it is worthwhile to stress the fact that documentation of Native American languages from older periods is scarce.

Historical Documentation of American Native Languages

The history of indigenous languages of the Americas had unfolded for many thousands of years before the earliest written records of them originated. The first known Amerindian community where the native language was written was that of the Olmecs in Mesoamerica. The central area of this oldest pre-Columbian civilization extended across the modern Mexican federal states of Guerrero, Veracruz, and Tabasco. The Olmec civilization developed the basic technologies and laid the foundations for institutions that were later adopted by the Mayans, Aztecs,

and other civilized peoples of the pre-Columbian era: writing, a calendrical system, monumental architecture, and so forth.

Longer texts in Olmec and, later, in Mayan date to the first millennium BCE and are contemporaneous with Greek and Roman literacy in Europe. In a comparative view, the written documentation of Native American languages is much younger than the tradition of writing in the Old World, where the beginnings lie with ancient Egypt in the fourth millennium BCE and with the Danube civilization in the fifth millennium BCE. The documentation of languages in the Americas over some 2,500 years is therefore much more limited than the written record of languages in Africa and Eurasia, and this has a bearing on the approaches to trace the splitting of individual languages, their branches, and whole language families (phyla) in the horizon of absolute time.

For the longest span of time in the history of Native American languages, no empirical evidence is available to identify the spread of languages and their splitting processes. What can be reconstructed with the methods of historical-comparative linguistics for prehistory are theoretical constructs, that is, fabrics of protolanguages whose real value as a means of communication remains questionable. Linguistics proper and anthropology played the role of forerunners for the study of languages in the Americas for more than a hundred years and well into the second half of the 20th century. During the past few decades, more and more insights into the formation of language families, about their historical relationships, and about the contacts involving their communities of speakers have been produced by interdisciplinary research. In addition to linguists and anthropologists, the American linguistic landscape has been studied by archaeologists, ethnologists, culture historians, and, more extensively since the 1990s, by human geneticists.

The documentation of Native American languages began in the 16th century. Following the model of the first grammar of a European vernacular language, Spanish, in 1492, European missionaries wrote the first grammars and compiled the first dictionaries of Amerindian languages. The first grammar was that of Tarascan, spoken in western Mexico, written by Maturino Gilberti and published in Mexico City in 1558. The classical Nahuatl language, the lingua franca of the Aztec Empire, was described by Alonso de Molina. This work—still of historical value—was printed in 1571. Many of the Spanish missionaries were interested in Amerindian languages, and the first valuable accounts about the number of individual languages come from the region of the Spanish colonies in America. Lorenzo Hervás y Panduro (1735–1809), a Spanish Jesuit, provided the first survey of languages in South America. In his universal encyclopedia—an edition with 21 volumes in Italian appeared between 1778 and 1787, a Spanish version in 6 volumes between 1800 and 1805—one finds much valuable information about American languages and their grammar and vocabulary.

Another remarkable source with collections of vocabularies from various American languages, the biggest enterprise of language studies during the age

of Enlightenment, is the monumental dictionary of Catherine II the Great, who ruled Russia from 1762 to 1796. The dictionary project (Vocabularium Catherinae) is associated with her name because she actively participated in the collection of its materials. As for the American languages, the czarina contacted representatives of American public life in personal letters, such as Benjamin Franklin and George Washington, to obtain data about indigenous languages. Benjamin Franklin, founder of the American Philosophical Society (1769), was perhaps the most knowledgeable American in matters of Amerindian cultures and languages at the time. Catherine's collections of linguistic data were organized by the German scholar Peter Simon Pallas in two volumes (1786–1789), published in Saint Petersburg. The collections of American languages are not included in this first edition of the *Linguarum totius orbis vocabularia comparativa*. The second, enlarged edition, which appeared in four volumes in 1790 and 1791, contains data from American languages.

In the early phase of data collection about American languages, the curiosity to explore the exotic world of indigenous cultures dominated over any systematic approach of their study. The two amateurs who published books about American languages in the 17th century—Roger Williams (*Key into the Language of America*, 1643) and John Eliot (*The Indian Grammar Begun: An Essay to Bring the Indian Language into Rules*, 1666)—had fanciful ideas about the relationship among individual languages. They thought that all Amerindian languages were more or less the same. Among the amateurs of the 18th century were the protagonists of the independence movement. Regna Darnell notes that Thomas Jefferson, "who devoted considerable energy to collecting Indian vocabularies in the years before his presidency," was one of these (Haarmann 2004: 780).

While in the context of language studies in Europe (especially relating to the Indo-European and Semitic languages) during the 18th century, knowledge was constantly accumulating about the historical relationships of language families, but the situation was much less promising with respect to languages in the Americas. The collections of linguistic data in the comparative enterprises of the 18th century, carried out by Europeans, did not yet allow a systematic approach to the genetic classification of American languages, although the encyclopedic work compiled by Hervás y Panduro is still of historical value. With an increase in the amount of data that became available about Amerindian languages in the course of the 19th century, reflections about their genetic classifications could be based on more solid grounds after that time.

The Linguistic Landscape of the Americas

Some 1,010 individual native languages are spread over the areas of the two Americas, the majority of them in South and Central America. This number is an approximation, because it gives an account of the present situation. The exact

number of Amerindian languages will never be known, because there is constant fluctuation. A hundred years ago there were more languages than today, and in a hundred years there will be fewer. The weakening of language maintenance in Amerindian speech communities and the loss of indigenous languages produced a marked trend of decline by the 19th century, and this trend has been accelerating in the past decades.

The decline of the vitality in Amerindian speech communities has been monitored for some 150 years. When Franz Boas (1858–1942), the German-born father of American anthropology, set out on his monumental enterprise to map out the ethnographic landscape of North America, he perceived the threat to the survival of many native languages. The loss of indigenous languages is a continuous process, and, in many cases, the date of extinction can be determined with the death of the last speaker. Examples of such processes are Omurano (Peru; extinct since 1958), Chumash (U.S./California; extinct since 1965), Jorá (Bolivia; extinct since 1963), Tillamook (U.S./Oregon; extinct since 1970), Yamana (Chile/Argentina; extinct since 1978), Nooksack (U.S./Washington; extinct since 1988), Twana (U.S./Washington; extinct since 1980), Yavitero (Venezuela; extinct since 1984).

Systematic surveys about the rate of loss of languages are available only for some regions such as for Brazil. Of the 230 Indian communities that still existed around 1900, altogether 86 had become extinct by 1950, either as a result of total assimilation of speakers of native languages to Portuguese or because the community became defunct with the overaging and the death of its members. Still in our times, there is uncertainty about the fate of certain speech communities. Despite the general trend of a loss of native languages and humanitarian concerns about the disintegration of world cultural heritage, the modern observer has to be cautious not to get entangled in a web of disinformation about the current situation. A number of languages and the communities of their speakers have been reported as extinct, but deeper investigations produced contradicting evidence of their survival. The language of the Karitiana in the Amazonian region was classified by Ribeiro (1957) as extinct, although living speakers have been reported in recent years. Similarly, other Amazonian languages such as Aruá, Mondé, or Arara (at the mouth of the Gi-Paraná) were listed as extinct in the 1960s but were "rediscovered" in the 1980s.

Pitfalls and Quandaries of Historical Classifications

The history of the classification of American languages illustrates that there are two aspects of the concept of "historical" classification. First, there is classification associated with the linguistic infrastructure. This relates to the taxonomies of historical reconstruction of genetic relationships between languages as applied by historical linguistics. Here, the focus is on the identification of the time depth of the splitting of individual languages from a common basis by means of

comparing cognate words and grammatical structures. Second, there is classification based on external factors that cause variation in the development of languages. External factors that shape the ecology of languages are manifold, and they may be environmental, social, cultural, or economic. The development of languages in contact may unfold under similar conditions of the landscape where they have spread (e.g., the Pueblo cultures in the arid zone in the U.S. Southwest). Certain social traditions may cause intense language contact and fusion of linguistic structures in certain areas (e.g., the custom of exogamy, that is marrying members of ethnic groups that are different from one's own, such as among many local communities in the Amazon region). The cultures of Amerindians (and their corresponding languages) may be (and have been) classified according to similarities of their socioeconomic traditions (e.g., grouping the Amerindian cultures in the U.S. Northeast according to the criterion of their common traditional system of subsistence: hunting in wooded terrain).

Historical classifications of American languages have been elaborated by focusing on both internal features (relating to the linguistic infrastructure) and external factors (ecological in a wider perspective). A general trend can be recognized in a retrospective of classification approaches. Purely linguistic taxonomies tend to produce a greater number of regional groupings (language families, or phyla) than those classifications that are more oriented at external factors of language ecology.

The first synopsis of northern Amerindian languages, based on observations of similarities (and dissimilarities, respectively) of lexical items, was accomplished by Albert Gallatin, Thomas Jefferson's secretary of the treasury, from a questionnaire he circulated starting in 1836. In his survey of 1848, Gallatin distinguished 32 language families. Information about Amerindian languages continually increased. A much more comprehensive survey of languages in the Americas originated in the late 19th century. In 1891 John Wesley Powell published his classification of Amerindian languages in which he distinguished altogether 55 independent stocks (later revised to 58). Powell had been director of the Bureau of Ethnology (under the auspices of the Smithsonian Institution, founded in 1846) since 1879 and had access to some 670 vocabularies that had been collected by the bureau. Powell's classification was mainly based on lexical comparisons and, given the lack of knowledge about historical sound changes at the time, it necessarily remained an analysis of surface value. Reliable information about the grammatical structures of Amerindian languages was still scarce at the time, so such data would not have sufficed for elaborating a survey.

Although deficient from the standpoint of modern linguistic taxonomy, Powell's classification, evaluated by himself as preliminary, nevertheless remained a yardstick for later conservative approaches to classification. Powell was concerned with the mapping out of local ecological conditions of neighboring speech

communities, while the overall currents of cultural history of the Americas played no role in his perspective. The orientation at a broader perspective, however, also found its reflection in enterprises to classify Amerindian languages. In 1921 Edward Sapir published his classification, aimed at the reconstruction of cultural history rather than at taxonomic perfectionism. Sapir distinguished only six "super stocks" with various subdivisions. Greenberg, with his fundamental distinction of only three macrophyla, reaches out for the extreme orientation at reconstructions of cultural history of the Americas. His radical position is marked by the special reconciliation of the parameters of his classification scheme with the insights about genetic fluctuations among the Amerindian populations.

John Wesley Powell, U.S. explorer and scientist. The Smithsonian Institution, under his directorship, published the first classification of Native American languages. (Library of Congress)

In a way, the classifications of Powell and of Greenberg mark positions on the extremes of a continuum of scientific parameters that are available for taxonomic purposes. It would be futile to compare Powell's 58 groupings with Greenberg's distinction of three macrophyla without any reservation. That would be something like comparing apples to pears. The two approaches differ greatly with respect to the architecture of the featural grid that was applied for each classification. It would be as futile to evaluate one survey as "right" and the other as "wrong," because the two classifications represent different levels of methodology.

There is another aspect to working with classifications that has to do with the hierarchical order of units of applied taxonomy. In the various classifications of Amerindian languages, different terminologies are applied. Key concepts such as *language family, subfamily, phylum, macrophylum, stock, super stock*, and others—not to speak of the multitude of subdivisions—are not synonyms but rather associate different meanings. Some terms are more comprehensive than others. For instance, one of the largest groupings of languages in South America is classified by some linguists as the "Tupí language family," with various regional subdivisions. Others are inclined to emphasize the greater independence of the subdivisions, which are themselves categorized as "language families" and as belonging to a "Tupí macrophylum." Classifying the

Tupí languages as one macrophylum or eight language families is a matter of the relativity of perspective, not one of being right or wrong.

Any discussion about the classification of languages is confronted with certain methodological limitations. Among the fields of science that are associated with language studies, some can explore deep layers of human evolution, while others do not reach far back in absolute time. Human genetics is in a privileged position, because its scientific methods favor investigations into the depth of prehistory to trace the fluctuation of populations from the early settlement of America onward. The situation of archaeology is less favorable. Through their excavations, archaeologists may retrieve artifacts along the trails of the great American migrations and the spread of people. Some types of artifacts (e.g., spear heads) are interpreted as leitmotifs, as typical markers of different stages of cultural development. And yet archaeology has to do with fragments of material culture, without any overall picture of human communities in prehistory.

Native American Migration

Scholars agree that at some point during the last Ice Age, a group of nomadic hunters crossed from eastern Siberia to Alaska by means of what was then a strip of land across the Bering Strait, perhaps pursuing the megafauna (mammoths etc.) that then still thrived. Some scholars hold that this group was responsible for all of the early settlement of the Americas, that they followed their game as far as South America, and that they were the progenitors of the Inca and Maya as well as the Inuits.

The Clovis culture, named for artifacts found in Clovis, New Mexico, has often been considered the culture descended from these land bridge crossers and was believed to have spanned much of the Americas. For decades, the lack of strong evidence of pre-Clovis settlements helped to support this theory.

In addition to the Bering land bridge migration, other scholars posit human migrations along water routes, with South America often believed to have been settled earlier than North America. Some models have Siberians traveling to the Northwest coast by boats, usually used for river travel (aided by the low sea levels of the era), Southeast Asians crossing the Pacific to South America, and Oceanic peoples crossing the Antarctic coast on their way to the South American tip. One Atlantic coastal migration model even suggested a Cro-Magnon progenitor for the Algonquians. Originally proposed in the 1930s, that Atlantic model has been largely displaced by the Solutrean hypothesis, proposed by Dennis Stanford and Bruce Bradley in 1999. According to their theory, the Solutrean people of prehistoric Europe crossed the Atlantic on small watercraft, with the aid of ice floes. These Solutreans would have been the progenitors of the Clovis culture, and an ancient site in Virginia is claimed as an example of a transition between the Solutrean and Clovis cultures as the people moved west.

Linguistic reconstructions of genetic relationships between languages are hampered by limitations of the time depth, which can be reached by taxonomic methodologies. This factor makes the linguistic identification of language families a rather tedious business.

The Time Depth of Historical Linguistics and the Chronology of the Splitting of American Language Families

There is an ongoing debate about the validity of methods as applied by historical linguistics to estimate the time depth for the emergence of language families. Methodology is the furthest advanced in the field of Indo-European studies. The individual languages of the Indo-European phylum are the best known and the most broadly investigated. There are more linguistic data available about Indo-European languages than about any other language family.

Despite the favorable conditions to study these languages that are genetically related, the methods to analyze the time depth of their splitting from a reconstructed protolanguage are not generally accepted. Widely applied are lexicostatistical methods. Structural differences of cognate expressions that are historically related (e.g., English *mother*, German *Mutter*, Latin *mater*, Russian *mat'*, Sanskrit and Old Persian *matar*, Old Irish *mathair*, etc.) in a variety of Indo-European languages are projected onto a time scale of assumed rates for sound change. The crucial issue with linguistic dating methods is the operation with average dates for language change. In modern research, the role of language contacts and linguistic interference in linguistic structures has been emphasized as a factor that blurs average dates of sound change and the splitting of a common basis into individual languages.

Some historical linguists accept an early date for the formation of a common protolanguage from which all known Indo-European languages derived, setting the date of its emergence at approximately 7000 BCE. Others are much more skeptical about the reliability of comparative methods, and their estimates are much more cautious, ranging between 4500 and 4000 BCE. The application of lexicostatistical methods that allow for the comparison of a large number of languages has produced a scaling of the time depth for the formation of many other language families, and the scalings are as controversial as for the Indo-European phylum.

As for the language phyla of the Americas, only rough estimates for the periods of their formation are as yet available. In a general overview, the following major groupings are to be distinguished:

1. Era of Formation between 4000 and 3500 BCE: Algonquian (northeastern U.S.; individual languages: Micmac, Cree, Ojibwa, Blackfoot, Cheyenne, etc.), Oto-Mangue (Central America, predominantly Mexico; individual languages: Chinantec, Mixtec, Zapotec, etc.).

2. Era of Formation between 3500 and 3000 BCE: Uto-Aztec (southwestern U.S., Mexiko, El Salvador; individual languages: classical Aztec, modern varieties of Nahuatl, Shoshone, Sonora, etc.).

3. Era of Formation between 3000 and 2500 BCE: Chibcha (southern regions of Central America, northern regions of South America; individual languages: Aruak, Chibcha, Kuna, Rama, etc.), Tupí (lowlands of Brazil, Paraguay; individual languages: Tupí, Guaraní, Wayampi, etc.), Panoan (Peru, north-eastern Bolivia, northwestern Brazil; individual languages: Capanahua, Nukuini, Karipuná, etc.)

4. Era of Formation between 2500 and 2000 BCE: Quechua (Andian region: Colombia, Ecuador, Peru, Bolivia, Chile; individual languages: Quechua of Cuzco, Ayacucho, Chimborazo, jungle Inga, etc.), Tucanoan (southern Colombia, Ecuador, Peru, Brazil; individual languages: Cubeo, Macuna, Secoya, etc.), Arawakan (Central America, northeastern regions of South America; individual languages: Guajiro, Carib, Taino, etc.), Mayan (southern Mexico, Guatemala, Belize, northwestern Honduras; classical Mayan languages are Chol and Quiché, modern languages are Tzeltala, Acatec, Ixil, Mopán, etc.).

5. Era of Formation between 2000 and 1500 BCE: Iroquoian (northeastern U.S.; Mohawk, Oneida, Seneca, etc.).

6. Era of Formation between 1500 and 1000 BCE: Athabascan (U.S./Alaska and southwestern regions; individual languages: Navaho, Carrier, Apache, etc.), Siouan (U.S./prairies of the Midwest; individual languages: Tutelo, Dakota, Crow, etc.), Mixe-Zoquean (Mexico/federal state of Oaxaca; individual languages: classical Olmec, Quetzaltepec, Popoluca, etc.).

The general impression of linguistic diversity in relation to the distinction of language families and their formation period is that even the maximum time depth of ca. 4000 BCE as reconstructed for Algonquian is fairly "shallow" when compared with the time depth of the early settlement of America. The span of time for which no historical-linguistic reconstructions are available ranges between 6,000 and 8,000 years, if not more. It must seem conclusive that the oldest dates for the formation of protolanguages come from North America, while those for language splittings in South America are considerably younger. These proportional differences in time scales can be reconciled with the migration movements of early settlers from north to south.

The time scales illustrate that, in the Americas, there are no old macrophyla like some found in Eurasia (i.e., the Indo-European, Uralic, or Afro-Asiatic language families, the protolanguages of which date to between 7000 and 8000 BCE). Even the oldest protolanguages that can be reconstructed for American phyla (i.e., the Algonquian and Oto-Mangue language families) are much younger. The era of the formation of protolanguages in America is comparable

to some of the younger groupings in Eurasia (e.g., the Austronesian, Dravidian, and Altajic language families).

An Interdisciplinary Approach to Classification

Whatever positions geneticists and diffusionists may take in the ongoing debate about Amerindian languages, no one can bridge the time gap between ca. 11000 BCE (the conventional date for the beginning of the peopling of America) and ca. 4000 BCE (the approximate date of the separation of the Algonquian phylum from an assumed common Amerind basis) with any linguistic reconstruction.

It cannot be reasonably ruled out that those Americans who created the earliest rock art spoke languages whose historical relationships were still recognizable as stemming from a common protolanguage. In general, pictures engraved or painted on rocks in the Americas have been produced from the end of the Ice Age—some 12,000 years ago into the historical era. It is noteworthy that the oldest dates attributed to rock art locations are found in South America, not in the Northern Hemisphere, although the Paleo-Indians arrived from the north. The controversial dates for sites such as Goias in northern Brazil (ca. 43,000 years ago) and Pedra Furada in northeastern Brazil (between ca. 30,000 and 25,000 years ago) do not fit the conventional time frame for the earliest migrations of humans from Siberia to North America (between 13,000 and 11,000 years ago). Taima-Taima in northwestern Venezuela (ca. 13,000 years ago) and Monte Verde in central Chile (ca. 12,500 years ago) still predate the oldest sites in North America.

Whatever the date of the first settlement and of the oldest rock pictures, the traces of the languages spoken in those remote times are lost. In the course of the migration movement from north to south, the process of a branching out of the protoform that may be considered the ancestor of all recent Amerind languages started sometime and unfolded until the oldest phyla can be identified by historical linguistics.

As long as there is no documentation or reliable reconstruction of the splitting of linguistic phyla in remote times, there is no way to disprove the validity of Greenberg's pan-American vision, albeit there has been substantiated criticism of methodological flaws in the comparative methods that were applied to identify historical relationships. Instead of trying to dislodge the overall perspective of a distribution of languages in accordance with the three-wave migration that is firmly anchored in the findings of archaeology and human genetics, it seems more reasonable to clarify the internal groupings of the Amerind macrophylum, for instance, by investigating with more scrutiny than hitherto whether Algonquian is Amerind or not.

Clarification of the linguistic interrelations within the Amerind complex is also needed. To this end, it is essential to establish a comprehensive catalog of grammatical and lexical equations that can also be acknowledged by the diffusionists.

There is a whole network of undeniably pan-American (Amerind) equations. An example is the pronoun system. The lexical roots to mark the first person (i.e., *n-*) and the second person (i.e., *m-*) are found in Amerind language from Canada to Chile. If directed to fruitful goals, the ongoing debate might even take on a special challenge for all of linguistics, and that is the task to refine historical-comparative methodology.

The search for the common origin of the Amerind languages, descendants of the language(s) of the first settlers who reached America, is no unreasonable quest, although, for the time being, such a remote protolanguage—although attempted—cannot be reliably reconstructed with the methods currently applied. And yet the connections between the languages of the second wave to America, the Na-Dene languages, and Eurasian languages (i.e., Palaeoasiatic, Caucasian) can be demonstrated with some certainty. The linguistic relationship of the languages of the third wave is undisputed, because the interconnections between the local varieties of Eskimo in America and Siberia can be evidenced beyond doubt.

It is interesting to draw a comparison with an earlier stage in comparative linguistics, at a time when Greenberg's unconventional classification of *The Languages of Africa* (1963) was first published. His novel ideas about how to systematize the genetic relationships of African languages stirred up vigorous opposition among linguists then as did his pan-American vision in the late 1980s, but it was equally valid. It is noteworthy that it did not even take two decades before Greenberg's classification of African languages became widely accepted. Here, too, the great currents of African history are reflected in the grouping of macro. In the case of the current controversy about American languages, it might be advantageous to write about it from a neutral standpoint (as a European linguist) and to view the linguistic landscape of America from an eagle's perspective in order to see the validity of Greenberg's hypotheses.

References and Further Reading

Bellwood, Peter. "The Time Depth of Major Language Families: An Archaeologist's Perspective." In *Time Depth in Historical Linguistics*, Vol. I. Edited by Colin Renfrew, April McMahon, and Larry Trask. Cambridge: McDonald Institute for Archaeological Research, 2000.

Bright, William. "Review of Language in the Americas (Greenberg 1987)." *American Reference Books* 23 (1988): 440.

Cavalli-Sforza, L. Paolo Menozzi Luca, and Alberto Piazza. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press, 1994.

Darnell, Regna. "Anthropological Linguistics: Early History in North America." In *International Encyclopaedia of Linguistics* (Vol. I, pp. 69–71). Edited by William Bright. New York: Oxford University Press, 1992.

Goddard, Ives. "Review of Language in the Americas (Greenberg 1987)." *Current Anthropology* 28 (1987): 656–57.

Greenberg, Joseph H. *Language in the Americas*. Stanford, CA: Stanford University Press, 1987.

Grimes, Joseph E., and Barbara F. Grimes. *Ethnologue Language Family Index*. Dallas: Summer Institute of Linguistics, 1996.

Haarmann, Harald. "Sociolinguistic Aspects of Cultural Anthropology." In *Sociolinguistics: An International Handbook of the Science of Language and Society*, 2nd ed. (pp. 769–85). Edited by Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier, and Peter Trudgill. Berlin: Walter de Gruyter, 2004.

Mithun, Marianne. *The Languages of Native North America*. Cambridge: Cambridge University Press, 1999.

Ribeiro, D. "Culturas e línguas indígenas do Brasil (Cultures and indigenous languages of Brazil)." *Educação e Ciencias Sociais* (Rio de Janeiro) 2 (1957): 1–102.

Ruhlen, Merritt. *On the Origin of Languages. Studies in Linguistic Taxonomy*. Stanford, CA: Stanford University Press, 1994.

Weltgeschichte der Sprachen. Von der Frühzeit des Menschen bis zur Gegenwart (World history of languages. From the times of earliest humans to the present). Munich: C. H. Beck, 2006.

CON

For well over 100 years, linguists, historians, and anthropologists have debated whether the diversity of Native American languages known to have prehistorically existed in the Americas can be reduced into just three grand linguistic roots, or whether the linguistic diversity is too great for such a reduction. The reasons behind this long-standing debate are complex and multifaceted, and over the years theories have been put forth that argue for one side or the other. Current empirical evidence indicates, however, that such a reduction is not possible and that Native American languages cannot be reduced beyond several dozen language roots, conclusively resolving the debate.

In fact, this is the only logical conclusion possible given the great linguistic diversity known to have prehistorically existed in the Americas, as well as the current physical anthropological and archaeological evidence. For example, over 300 distinct, mutually unintelligible languages are known to have been spoken north of the Rio Grande River before the arrival of Europeans. Likewise, over 1,500 languages were spoken in Central and South America before the arrival of Europeans; countless others disappeared before any documentation of their existence could be completed. Furthermore, the ones that we do know about and that

have been adequately studied differ in fascinating ways not only from the better-known languages of Europe and Asia, but also among themselves in their sounds, in the concepts they package into words, in their grammatical categories, and in their larger patterns of expression. While it has been possible to classify the languages of Europe into just three roots—Indo-European, Finno-Ugric, and Basque—similar attempts to reduce the vast diversity of Native American languages into a similar tripartite classification scheme have been met with skepticism.

Attempts to Find Roots

Original attempts to find the linguistic roots of Native American languages began early in the 20th century with the work of Paul Radin (1919) and Edward Sapir (1912, 1917). However, it was not until James Greenberg (1987) published his highly contentious linguistic consolidation theory that the languages of Native Americans were reduced to just three grand linguistic roots. Although most linguists rejected this proposition, some physical anthropologists accepted it, as it seemed to fit with their data. For example, G. Richard Scott and Christy Turner II (2000) argued that such a reduction corroborated their evidence based on dental morphological characteristics of Native Americans and north and south Asians. Similarly, S. L. Zegura (1984) argued that this hypothetical tripartite linguistic classification scheme fit with early results obtained from genetic studies. However, as linguists examined Greenberg's theory and further evidence was gathered in other fields concerning the early prehistory of the Americas, this tripartite linguistic classification scheme became untenable. Rather, current empirical evidence conclusively demonstrates that it is impossible to reduce the languages of Native Americans beyond a few dozen linguistic roots.

Edward Sapir

In addition to his work with Radin and on the Athabaskan languages in general, Edward Sapir (1884–1939) was one of the most influential and accomplished linguists of his generation. Much of his work continues to polarize linguists in the 21st century, as the theories he set out formed a significant portion of early structural linguistics. In his landmark study *Language: An Introduction to the Study of Speech* (1921), he put forth an extensive grammar-typology, classifying the languages of the world according to their grammatical structures. Most famously, some of his ideas were incorporated by his colleague Benjamin Whorf into the Sapir-Whorf hypothesis (Whorf 1949), which states that the nature and structure of a person's native language determines (or put more weakly, affects) that person's experience of and interaction with the world. The Sapir-Whorf hypothesis has come under a great deal of criticism in the past few decades, and the debate has been one of the most important in linguistics and cognitive science.

Between the beginning of the 20th century and the present day, linguistic methods for reconstructing language affiliations through time have grown in rigor and sophistication. The first to classify Native American languages in some continental fashion was John Wesley Powell (1891a, 1891b). Powell's classification of languages north of Mexico included 58 roots (or "families") and became the baseline for subsequent work in the classification of Native American languages. As Sapir later expressed during his own work on classifying Native American languages, "the cornerstone of the linguistic edifice in aboriginal North America" was provided by the early work of Powell (Sapir 1917: 79). Before this, Spanish and French colonialists and early Euro-Americans such as Roger Williams (1973), Peter Stephen Duponceau (1819a, 1819b), Albert Gallatin (1973), and Robert Gordon Latham (1845, 1856) did work in the languages of Native America, but none had attempted to classify all the languages of the Americas on such a geographic scale.

Because Powell was the first to attempt a classification of Native American languages, the method he used was not very developed. Instead, it was a rather impressionistic inspection of rough word lists and vocabularies gathered from early encounters between Euro-Americans and Native Americans. According to Powell, "The evidence of cognation [that languages are derived from a common ancestral family] is derived exclusively from the vocabulary" (1891a: 11). Franz Boas (1911) subsequently took up the task of refining the linguistic understanding of Native American languages, building on the work of Powell. As has been well documented in the writings of Regna Darnell (1969) and Andrew Orta (2004), Boas came to be associated with a cultural particularist approach to language and culture, in which he compared and contrasted the typological traits of languages in a particular geographic area to determine how they might have been reshaped as a result of mutual influence in that limited area. Out of this work, Boas documented how difficult it was to distinguish linguistic traits that were the result of a genetic linguistic relationship from those that were a result of simple linguistic borrowing or cultural processes. Because of this work, Boas cautioned those attempting to reduce the diversity of Native American languages into only a few grand linguistic roots, because determining linguistic affiliation was extremely difficult.

Others working with Native American languages at this time, such as Alfred Kroeber (1909), Edward Sapir, and Paul Radin, also published influential work on the debate of how many linguistic roots were present in the Americas. Kroeber, for example, worked on the languages of California and the Great Basin, while Sapir worked on Plateau languages; both worked on historical-linguistic affiliations. Kroeber was not in favor of reducing the linguistic diversity of Native American languages beyond specific geographic areas, while Sapir became known as a strong advocate of distant hypothetical linguistic roots that combined all Native American languages. Radin, conversely, focused primarily on merging
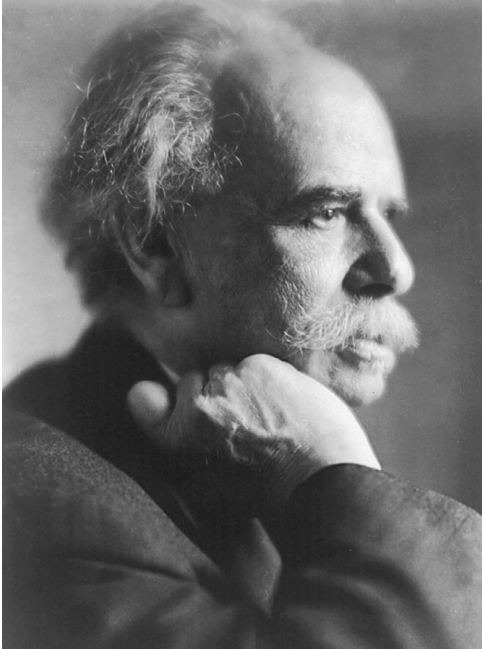
Franz Boas, who helped spread the discipline of anthropology in the United States, was a scholar of broad learning, concerned with all aspects of humans and human culture. (Library of Congress)

all known languages into just a few roots rather than investigating specific languages or linguistic patterns of change. He argued that all Native American languages were genetically related and belonged to one large linguistic root. He saw in his colleagues' work (that of Alfred Kroeber and Edward Sapir) only 12 remaining independent roots and believed that merging them into one was "hardly so revolutionary." However, according to Lyle Campbell, most of Radin's contemporaries did not accept his attempt to unite all these languages, primarily because the evidence for such a unification was highly conjectural.

Today, Native American language classification has been greatly influenced by the opinions of these early linguists. As a result of these historical attempts to understand and document the relationships of Native American languages, two schools of thought developed over time, resulting in today's debate and its opposing sides. Those that followed the consolidation process of Sapir, Radin, and others continued to look for evidence that Native American languages could be reduced to just a few grand linguistic roots, while followers of Boas, Kroeber, and others maintained that such a reduction was overly simplistic.

Greenberg's Consolidation Theory

Over time the debate came to a standstill, because little evidence to resolve it was forthcoming and because linguists concerned themselves with other aspects of Native American languages. In 1987, however, Greenberg revived the consolidation theory with the publication of *Language in the Americas*, sparking renewed interest in the debate. In this book Greenberg argued that it was possible to reduce all Native American languages into just three grand linguistic roots, each of which represented a unique migration wave separated in both space and time. These three grand linguistic roots, dubbed Amerind, Na-Dene, and Eskimo-Aleut, were argued to represent the linguistic roots of all Native American languages and claimed to be the culminating results of Sapir's methods. As both

Victor Golla (1988) and Robert Rankin (1992) have independently pointed out, however, the methods of Greenberg and Sapir are fundamentally different, in spite of their shared interest in large-scale consolidation of linguistic roots.

A basic fact on which all linguists agree concerning Native American languages is that historically there was extensive linguistic diversity in the Americas and that within this diversity, various levels of inclusively existed, resulting in linguistic roots. Greenberg, however, went beyond this general consensus and claimed that the Americas were settled by three separate population movements, each of which contained a different linguistic root for a total of just three in the Americas. Historical linguistics, however, as Campbell (1997) and Mithun (1990) have noted, is only able to reliably reduce the diversity of Native American languages to approximately 55 genetic roots in North America, 10 in Central America, and more than 80 in South America, totaling approximately 150 distinct linguistic roots for all of the Americas. One of the central components behind this long-standing debate is the confusion stemming from the terminology used to argue each side. To clarify my side of the argument, it is important to briefly discuss some of the terms used within linguistics and the debate.

Linguistics and Debate Terms

The term "dialect" is generally used to mean only a variety (regional or social) of a language that is mutually intelligible with other dialects of the same language. "Language," conversely, means any distinct linguistic entity that is mutually unintelligible with other languages. A language *root* is a group of genetically related languages, ones that share a linguistic kinship by virtue of having developed from a common earlier ancestor. Thus, it is common to find linguistic roots being designated with the suffix *-an* (e.g., Algonquian, Athabascan, Uto-Aztecan). Furthermore, it is important to note that language roots can be of different magnitudes. That is, they can have different time depths, with some larger-scale roots including smaller-scale roots as their members or branches (e.g., Celtic is a language root that has a shallower time depth than the larger language root of Indo-European, of which Celtic is part). Within this basic terminological structure, linguists have used a wide array of confusing terms to distinguish more inclusive from less inclusive roots. For example, the term "subroot" (also termed "subgroup" or "branch") refers to a group of languages within a well-defined language root that is more closely related to each other than to other languages of that root; they constitute a branch of the phylogenetic tree of the language root (i.e., Numic is a subroot of the larger Uto-Aztecan language root).

Terms that have been used for postulated but undemonstrated higher order, more inclusive roots such as in the present debate (i.e., proposed distant genetic relationships) include stock, phylum, and the compounding element macro- (as in macroroot, macrostock, and macrophylum). These terms have become

confusing and controversial, as might be expected when proposed names for entities that are not fully agreed to exist are at stake (such as Greenberg's Amerind). Stock is ambiguous in that in older linguistic usage it was equivalent to language family (a direct transfer of the common German linguistic term *Stamm* [or *Sprachstamm*]). However, the term has often been used in America to indicate a postulated but unconfirmed larger long-range grouping that would include more than one established language root or genetic unit, such as William Shipley's (1980) use of the proposal of macro-Penutian in the Plateau region of North America. Finally, the terms phylum and macro have also been used to designate large-scale or long-range proposed but unestablished language roots.

To avoid any misunderstandings in the argument being presented here, these terms will not be used, instead the term root is solely used, because it appears both sufficient and not as controversial. This is because if the entities called stock, phylum, or macro were found to be correct, they would in fact be roots. Therefore, such terms are spurious, and it is more parsimonious to simply refer to these proposed higher orders more inclusive groups as hypothetical linguistic roots.

Lack of Methodological Agreement

The tripartite linguistic classification of Native American languages (that is, Eskimo-Aleut, Na-Dene, and Amerind) is not new, but reflects the culmination of reductionistic methodological processes that began at the start of the 20th century. Beyond the general confusion surrounding the terminology of the debate, a more substantial critique for not agreeing with the tripartite linguistic classification is the fact that there is no agreed method of reconstructing linguistic genetic affiliations at a deep time scale. In fact, there is no agreed method of chronologically determining when various languages and language roots diverged from each other. This is particularly true when attempting to unite large geographic areas or reconstruct languages beyond a few thousand years such as in the Americas. In fact, Peter Forster and Alfred Toth have convincingly demonstrated that even within linguistic studies of Indo-European, the largest and best-documented language root in the world, "the reconstruction of the Indo-European [phylogenetic] tree, first proposed in 1863, has remained controversial" (2003: 9079). Furthermore, unlike languages from Europe or other parts of the world, Native American languages have no tradition of older written texts on which a study of their history can be based. This has resulted in skepticism concerning any conclusions reached in the study of Native American languages because the linguistic deductions are thought to not be as sound as those from other parts of the world. However, as Ives Goddard argued, just because "documents and documentation are rarely accorded the attention they receive in the traditional study of Old World languages" (1973: 728) does not mean that the conclusions reached by careful historical-linguistic work are of a

spurious nature. Rather, because of this fact, historical linguistics has had to develop a sophisticated methodology for investigating language change and linguistic affiliation across space and time.

Methods to Investigate Linguistic Affiliation in the Americas

Two primary methods have been used when investigating the question of linguistic affiliation in the Americas: historical linguistics and multilateral word comparison. The method of historical linguistics is widely used and has been stringently developed for the past hundred plus years, ever since it was originally proposed by August Schleicher (1983). Contrary to this, the method of multilateral word comparison was developed only recently by Greenberg, who adapted principles from the glottochronology of Morris Swadesh et al. (1954) and Sapir's lexical, morphological, and phonological comparative method. According to M. Ruhlen (1986, 1987), in the method of multilateral word comparison, lists of words from the different languages under comparison are generated based on superficial similarities of sound and meaning, along with discursive considerations of similarities in grammatical morphemes. The primary aim of the method is classification, but the classification that results from it is simply a codified statement of the judgments of similarity that have been made in assembling the sets of words across the languages under comparison. Golla calls this method "the inspectional route to genetic classification" (1988: 93), while Calvert Watkins calls it "etymology by inspection" (1990: 293). The terms used by Golla and Watkins reflect the fact that this method depends essentially on lexical similarities determined largely by visual inspection.

The historical-linguistic approach, in contrast, dubbed "the major alternative" by Greenberg and colleagues (1986), employs standard techniques of historical linguistics to attempt to work out the linguistic history of the languages involved. Further, unlike the multilateral word comparison method, the approach of standard historical linguistics employs techniques for formulating and testing hypotheses about the undocumented history of languages. These techniques have been developed and refined over the past century on the basis of the study of the historical changes undergone by a wide variety of languages. The goal of historical linguistics, therefore, is to determine the principles and factors that govern language change. Once the principles of language change have been determined, it then becomes possible to investigate affiliations between languages across space and time.

Of primary concern here is the fact that after related languages have been separated for only a few thousand years, the resemblances between them resulting from their historical connections decrease through normal linguistic changes. The longer languages have been separated, the harder it becomes to develop a proper phylogenetic tree demonstrating the history of the particular language and

how it relates to other languages. This is why it is important to follow the historical-linguistic approach of establishing principles of language change. Because the multilateral word comparison method skips this step, however, it suffers from potentially biased data through the comparison of words from possibly nonsynchonous languages.

This central limitation of the multilateral word comparison method is particularly evident in the data used to support arguments for the tripartite linguistic classification scheme. For example, the data Greenberg used to support his argument were of poor quality, often drawn from brief early notes made by explorers passing through an area for the first time rather than the rich, linguistically superior dictionaries and grammars now available for many languages. Furthermore, as discussed by Mithun, in an attempt to increase the compatibility of the lists generated through the multilateral word comparison method, Greenberg retranscribed many of the lexical items into his own phonetic system, apparently without knowledge of the actual phonetic systems of the languages under comparison. Thus, numerous errors were introduced into Greenberg's dataset, and the retranscription process used by Greenberg rendered it impossible to recover the original sources of the material, none of which were cited, because "listing all these sources in a general bibliography would have added greatly to the length and cost of the work" (Greenberg 1987: xv).

Not only is the method of multilateral word comparison of a dubious nature, but some of the languages used as data for comparison by Greenberg are also of a spurious nature. For example, Greenberg introduced some language names into his dataset that are not languages at all. Membreno, which Greenberg classified as a Lencan language from Central America, is actually the name of a person (Alberto Membreno), whose work contains several Lencan word lists from different Honduran towns. Similarly, in several instances Greenberg gave the names of towns where a certain language was spoken as names of distinct languages. For example, Greenberg lists six Lencan languages when there are currently only two known; the other four are towns where Lencan is spoken. Although these errors are unlikely to greatly affect the overall tripartite linguistic classification scheme developed by Greenberg, they do indicate that the tripartite linguistic classification is highly conjectural and rests on unsound evidence. Furthermore, while it is generally agreed that basic vocabulary is, on the whole, more resistant to replacement than lexical items from other sectors of the vocabulary, such basic words are, in fact, also often replaced, so that even in clearly related languages, not all basic vocabulary reflects true cognates. This was one of the valid insights of Swadesh's glottochronology, generally discredited as a method of dating, but nevertheless based on the valid observation that even basic vocabulary can be and is replaced over time. This is a fundamental problem in attempting to reconstruct languages and their linguistic roots far back in time.

Greenberg acknowledges that if Native American immigrants left no linguistic relatives in Asia and died out in the Americas, there would be no linguistic evidence of their presence in the Americas. It is quite possible that groups, particularly if they were small, speaking a particular language or dialect may have simply died off, and consequently there might be no linguistic connection between America's earliest colonists and contemporary Native Americans or indigenous Asian groups, complicating any attempts at reducing the languages of Native Americans to just three grand linguistic roots.

Empirical Evidence Argues against Language Root Reduction to Three

Not only does the linguistic evidence not support the reduction of Native American languages into just three grand linguistic roots, but as mentioned earlier, empirical evidence from other fields also argues against such a reduction. For example, the molecular genetic evidence, based on haplogroup frequencies of genetic markers found on both the mitochondrial DNA (mtDNA) and the Y chromosome, argue that early Native Americans originated from a broad geographic area in Asia, and that this area does not correspond to any of the linguistic roots proposed by Greenberg and those in favor of the tripartite linguistic classification scheme. The molecular genetic data indicate that the initial migration into the Americas originated in south-central Siberia between 35,000 and 20,000 years before the present. These early migrants are hypothesized to have followed the Northwest coast route until they were south of the glacial ice sheets, where they expanded into all continental regions and brought with them mtDNA haplogroups A–D and Y chromosome haplogroup P-M45a and Q-242/Q-M3 haplotypes.

The molecular genetic evidence further indicates that a later migration entered the Americas, bringing mtDNA haplogroup X and Y chromosome haplogroups P-M45b, C-M130, and R1a1-M17, possibly using an interior route. Because these haplogroup markers come from a wide area in Asia, and because the Asian languages represented within these areas cannot be reduced into just three linguistic roots, it is argued that early Native Americans had a wider linguistic base than just three grand linguistic roots. Other physical anthropological data, such as craniomorphology and dental morphology, also support the conclusion that early Native Americans came from a wider geographic area that included several dozen linguistic roots and that attempts to reduce the linguistic roots into just a few linguistic roots has not been possible for the area.

Conclusion

There is great linguistic diversity in the Americas. While some scholars debate how many linguistic roots Native American languages can be reduced to, most believe that there are approximately 150 different language roots in the

Americas that cannot be shown to be related to each other. In spite of this diversity, it is a common hope that future research will be able to demonstrate additional genetic relationships among some of these roots, reducing the ultimate number of genetic units that must be recognized. However, the linguistic diversity that currently must be acknowledged means that on the basis of language classification, as well as other empirical evidence, we are unable to reduce the diversity of languages in the Americas to just three grand linguistic roots.

References and Further Reading

Boas, F. *Handbook of American Indian Languages*, Vol. bulletin IX. Washington DC: Government Printing Office, 1911.

Bright, William. "On Linguistic Unrelatedness." *International Journal of American Linguistics*. 36 (1970): 288–90.

Campbell, Lyle. *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press, 1997.

Darnell, Regna. "The Development of American Anthropology, 1879–1920: From the Bureau of American Ethnology to Franz Boas." Dissertation, University of Pennsylvania, 1969.

Duponceau, Peter Stephen. "A Correspondence between the Rev. John Heckewelder, of Bethlehem, and Peter S. Duponceau Esq., Corresponding Secretary of the Historical and Literary Committee of the American Philosophical Society, Respecting the Languages of the American Indians." Transactions of the Historical and Literary Committee of the American Philosophical Society, Vol. I, pp. 351–465, 1819a.

Duponceau, Peter Stephen. "Report of the Corresponding Secretary to the Committee, of His Progress in the Investigation Committed to Him of the General Character and Forms of the Languages of the American Indians." Transactions of the Historical and Literary Committee of the American Philosophical Society, Vol. I, pp. 14–46, 1819b.

Forster, Peter, and Alfred Toth. "Toward a Phylogenetic Chronology of Ancient Gaulish, Celtic, and Indo-European." *Proceedings of the National Academy of Sciences* 100, 15 (2003): 9079–84.

Gallatin, Albert. *A Synopsis of the Indian Tribes within the United States East of the Rocky Mountains and in the British and Russian Possessions in North America*. New York: AMS Press, 1973/1836.

Goddard, Ives. "Philological Approaches to the Study of North American Indian Languages: Documents and Documentations." In *Linguistics in North America* (pp. 727–45). Edited by T. A. Sebeok. The Hague: Mouton, 1973.

Goddard, Ives, and Lyle Campbell. "The History and Classification of American Indian Languages: What Are the Implications for the Peopling of the Americas?" In *Method and Theory for Investigating the Peopling of the Americas* (pp. 189–207). Edited by R. Bonnichsen and D. G. Steele. Corvallis, OR: Center for the Study of the First Americans, 1994.

Golla, Victor K. "Review of *Language in the Americas* by Joseph H. Greenberg." *American Anthropologist* 90 (1988): 434–35.

Greenberg, James. *Language in the Americas*. Stanford, CA: Stanford University Press, 1987.

Greenberg, J., C. G. Turner II, and S. Zegura. "The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence." *Current Anthropology* 27 (1986): 477–97.

Jones, Peter N. *American Indian mtDNA, Y Chromosome Genetic Data, and the Peopling of North America*. Boulder, CO: Bauu Institute Press, 2004.

Kroeber, Alfred Louis. "Paiute." *Notes on Shoshonean Dialects of Southern California* (pp. 256–62). Edited by A. L. Kroeber. Publications in American Archaeology and Ethnology, Vol. VIII. Berkeley: University of California Press, 1909.

Latham, Robert G. "Miscellaneous Contributions to the Ethnography of North America." *Transactions of the Philological Society* 2 (1845): 31–50.

Latham, Robert G. "On the Languages of Northern, Western, and Central America." *Transactions of the Philological Society* 14 (1856): 57–115.

Lightner, Theodore M. "On Swadesh and Voegelin's 'A Problem in Phonological Alternation'." *International Journal of American Linguistics* 37 (1971): 227–37.

Mithun, Marianne. "Studies of North American Indian Languages." *Annual Review of Anthropology* 19 (1990): 309–30.

Orta, Andrew. "The Promise of Particularism and the Theology of Culture: Limits and Lessons of 'Neo-Boasianism'." *American Anthropologist* 106, 3 (2004): 309–30.

Powell, John Wesley. "Indian Linguistic Families of America North of Mexico." In *Seventh Annual Report, Bureau of American Ethnology* (pp. 1–142). Washington DC: Government Printing Office, 1891a.

Powell, John Wesley. "The Study of Indian Languages." *Science* 17, 418 (1891b): 71–74.

Radin, Paul. *The Genetic Relationship of the North American Indian Languages*, Vol. XIV. Berkeley, CA: University of California Press, 1919.

Rankin, Robert L. "Review of *Language in the Americas* by Joseph H. Greenberg." *International Journal of American Linguistics* 58 (1992): 324–51.

Ruhlen, M. *A Guide to the World's Languages*, Vol. I. Stanford, CA: Stanford University Press, 1986.

Ruhlen, M. "Voices from the Past." *Natural History* 96, 3 (1987): 6–10.

Sapir, Edward. *Culture, Language and Personality*. Berkeley: University of California Press, 1949.

Sapir, Edward. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Company, 1921.

Sapir, Edward. "Language and Environment." *American Anthropologist* 14, 2 (1912): 226–42.

Sapir, Edward. "Linguistic Publications of the Bureau of American Ethnology, a General View." *International Journal of American Linguistics* 1 (1917): 280–90.

Schleicher, August. "The Darwinian Theory and the Science of Language." In *Linguistic and Evolutionary Theory: Three Essays by August Schleicher, Ernst Haeckel, and Wilhelm Bleek* (pp. 1–70). Edited by K. Koerner. Amsterdam: John Benjamins, 1983/1863.

Scott, G. Richard and Christy Turner II. *The Anthropology of Modern Human Teeth: Dental Morphology and its Variation in Recent Human Populations*. New York: Cambridge University Press, 2000.

Sherzer, Joel. *An Areal-Typological Study of American Indian Languages North of Mexico*. Amsterdam: North-Holland, 1976.

Shipley, William. "Penutian among the Ruins: A Personal Assessment." *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 11 (1980): 437–41.

Swadesh, Morris, et al. "Symposium: Time Depths of American Linguistic Groupings." *American Anthropologist* 56, 3 (1954): 361–77.

Watkins, Calvert. "Etymologies, Equations, and Comparanda: Types and Values, and Criteria for Judgement." In *Linguistic Change and Reconstruction Methodology*. Edited by P. Baldi. Berlin: Mouton de Gruyter, 1990.

Whorf, Benjamin L. *Four Articles on Metalinguistics*. Washington, DC: Foreign Service Institute, Department of State, 1949.

Williams, Roger. *Key into the Language of America*. Detroit: Wayne State University Press, 1973/1643.

Zegura, S. L. "The Initial Peopling of the Americas: An Overview from the Perspective of Physical Anthropology." *Acta Anthropogenet* 8, 1–2 (1984): 1–21.

15

The historical Buddha was born in 563 BCE and lived to 483 BCE.

PRO   Anita Sharma
CON K. T. S. Sarao

PRO

Several methods have been employed by historians, archaeologists, and astrologers, among others, to calculate the lifespan of Buddha. But most of them are unreliable, especially those that either depend on very late materials or are of a dubious nature. The literary and archaeological source material available for the construction of ancient India's exact historical chronology is totally insufficient and unsatisfactory, and thus almost all the dates are quite tentative. However, we can look at the historical events surrounding Buddha's life and determine that the traditional dates of Buddha's life, 563–483 BCE, are valid. Buddha's dates are inextricably linked to the date of King Ashoka's accession. There are four reasons for this. First, no concrete date prior to Ashoka is available in ancient India, with the exception of the invasion of Alexander (327–326 BCE) and the beginning of Chandragupta Maurya's reign (calculated variously between ca. 321–313 BCE). Chandragupta, who began his reign a few years after Alexander's invasion, was Ashoka's grandfather and, according to the *Puranas* and the Sri Lankan chronicles, began his reign 49 years before Ashoka did. Second, almost all the textual sources that provide information relating to the date of the Buddha use Ashoka as a reference point. Third, according to A. K. Narain in *The Date of the Historical Śākyamuni Buddha*, both of the Buddhist sectarian traditions "are interested, and compete, in 'possessing' Aśoka in relation to the date of the Buddha *without* fixing a definite date for the latter first" (1994: 43). Fourth, Ashoka is the earliest historical personality who is intimately connected to Buddhism and provides epigraphical (thus more reliable than literary) information on the Buddha, including his birthplace, different holy books, and teachings. Thus, Ashoka appears to hold the key, and his lifespan needs to be determined first before any assumptions can be made on the lifespan of the Buddha.

On the basis of the names of various Greek kings (Antiochus, Ptolemy, Antigonas, Magas, and Alexander) mentioned in Rock Edicts II and XIII, the date of Ashoka's accession may be put at about 268 BCE and the coronation (*abhisheka*), which took place in the fourth year of his reign (that is, after three

325

years), in about 265 BCE. These dates for Ashoka have been accepted more or less satisfactorily by modern scholarship.

Theories on Dating Buddha

The different hypotheses relating to the calculation of Buddha's dates fall within two categories that are based on ecclesiastical division in Buddhism: the northern Buddhist tradition and the southern Buddhist tradition. The calculations of the northern Buddhist tradition place the date of the *Mahaparinirvana* of the Buddha about 100 or 110 years before the coronation of Ashoka. The leading proponents of this date include H. Bechert (1982, 1991–92, 1995), K. T. S. Sarao (1989), and R. Gombrich (1991). This form of calculation, also known as short chronology, has been criticized on the ground that it appears to be "a fabrication by the monks of the Sarvāstivāda sect of Mathurā, who wanted to connect Upagupta with Aśoka. As a result of this device, the coronation year of Aśoka was moved to fit in with the years of Upagupta's life, i.e. around 100 years after the Nirvāna" (Yamazaki 2003: 147). This short chronology also fails to do justice by adjusting the periods of the Vinaya elders (who were guardians of the *Vinaya*) as well as the large number of kings who ruled in India and Sri Lanka between the period of the Buddha's time and that of King Ashoka.

Here we are concerned with the traditionally accepted date of about 483 BCE, the calculation of which is primarily done on the basis of the southern Buddhist tradition. The southern Buddhist legends contained in the *Dipavamsa*, the *Mahavamsa*, and the Pali version of the *Samantapasadika* place the consecration of Ashoka 218 years after the *Mahaparinirvana* (the Great Decease) of the Buddha. The best survey of the arguments that lead scholars to believe the calculation of Buddha's dates should be based on what is called the long chronology as found in Andre Bareau's research paper, "La date du nirvāna," published in 1953. The southern Buddhists had in the beginning adopted 544–543 BCE as Buddha's year of death. But this was later recalibrated by Geiger (1912) and others, who pointed out that 60 years extra had been added into the chronology of the kings of Sri Lanka and thus there was the need for a recalibration. One of the main arguments for the validity of Geiger's chronological calculations was a theory proposed by D. M. Z. Wickremasinghe (1904–12) in "Kiribat-Vehera Pillar Inscription" that a chronology starting from 483 BCE as Buddha's death date was known and had been used in Sri Lanka until the beginning of the 11th century and that the *Buddhavarsha* of 544 BCE was generally accepted at a later date. Indications are to be found that in earlier times, and indeed down to the beginning of the 11th century, an era persisted even in Ceylon in which 483 was reckoned as Buddha's death year. From the middle of the 11th century, the death year was presumed as 544 BCE, and this date is still in use.

As to Parākramabāhu I, we have information from inscriptions, confirmed and completed by literary data, that he was crowned 1,696 years after the Buddha's death, that is, in the year 1697 AB (after Buddha). Eight years later, in 1705 AB, a second coronation apparently took place. In the fourth year after that, when 1,708 years had gone by since the *Nirvāna*, that is, in 1709 AB, he held a Buddhist synod. According to the Ceylonese era, those are the years 1153, 1161, and 1165 CE (Ceylonese era). But this date for Parākramabāhu is supported by an entirely independent source, namely a South Indian inscription at the Temple of Tiruvālīśvara in Arpākkama.

According to Culavamsa, the six predecessors of Parākramabāhu, from Parākrama Pāndu onward, reigned for 107 years. Thus the accession of the last-named prince falls at 1590 AB, or according to the Ceylonese era, 1046 CE. Moreover, this date is confirmed by the south Indian *Manimangalam* inscription, which is dated in the same year. According to the latter, Parākrama Pāndu was conquered and killed in this year by the Cola king Rājādhirāja I. It is true that the Culavamsa gives Parākrama Pāndu a reign of two years, but we must rather take the accession and death of the king as falling in one and the same year, 1590 AB, that is 1046 CE.

But a date for Udaya III among the predecessors of Parākrama Pāndu can also be fixed from a south Indian inscription, which throws a completely new light on the whole reckoning of eras. Since, according to the Culavamsa, the time between the accession of Udaya III and that of Parākrama Pāndu amounts to 93 years and 8 days, and the latter ascended the throne in 1590 AB, we consequently have the date 1497 AB for the accession of this former king. But this year, according to the Tanjore inscription of King Rajendra Coladeva, must be about the year 1015 CE. The inscription gives an account of a military expedition to Ceylon. The details of the invasion by Cola correspond with one that occurred under Udaya III at the beginning of his reign. Udaya III's expedition falls between the fourth and sixth years of the reign, that is, between 1015 and 1018. This year must coincide with the years 1497 and 1498 AB. The difference between 1,497 years and 1,015 equals 482, which falls within pre-Christian times. This would mean Buddha died in 483 BCE.

Foundation of Long Chronology

The cornerstone of the long chronology is the number 218 mentioned repeatedly in the Sri Lankan chronicles: the *Dipavamsa* and the *Mahavamsa*. For instance, the *Dipavamsa* (VI.1) in this regard says, "218 years after the Sambuddha had attained Parinirvana, Piyadassana [Ashoka] was consecrated." Similarly, the *Mahavamsa* (V.21) goes on to say, "After the Conqueror's Nirvāna and before his [Aśoka's] consecration there were 218 years, this should be known." We are also told in the *Dipavamsa* and the *Mahavamsa* that the unrest that led to

the Third Council arose at the Asokarama in Pataliputta 236 years after the death of the Buddha, and that this council was completed in the 17th year of Ashoka's reign. This also places the *Mahaparinirvana* 218 years before the consecration of Ashoka.

Since the date of Ashoka's accession is calculated to about 268 BCE, the Buddha's death may be computed to about 486 (268 plus 218) BCE. But if the three years above are not accounted for, the death of the Buddha is put in the year (218 plus 265) about 483 BCE. It may also be interesting to note that the length of Bindusara's reign in the Sri Lankan tradition is given as 28 years, as against 25 of the *Puranas*. In all probability, this was because the Sri Lankan tradition included the three years before Ashoka's consecration in the reign of Bindusara. But the total length of Ashoka's reign was not changed likewise in these records.

As pointed out by Andre Bareau, the *History of Khotan* places the start of King Ashoka's reign in year 234 of the Buddhist era (BE), which is not very different from the long chronology's 218 BE. The dates of the long chronology also appear to be supported by the events of contemporary political history. For instance, the lists of Magadhan kings in different sources, though showing discrepancies on many points, are nevertheless unanimous in placing several kings between Ajatashatru and Chandragupta, the grandfather of King Ashoka. These lists can only be adjusted satisfactorily between the Buddha and Ashoka by following the long chronology. Another important reason the long chronology appears to be more logical is that, instead of the suspicious number of 100 in the short chronology, the long chronology has the exact number of 218.

W. Geiger's discussion of the chronology of the Buddha in his *The Mahāvamsa or the Great Chronicle of Ceylon* played an extremely important role in getting acceptance for the long chronology as against the short chronology. Other scholars like Andre Bareau and P. H. L. Eggermont (1956, 1969) also followed suit, and thus the long chronology became the basis for the date of the Buddha. However, the biggest landmark that provided justification for the long chronology came in the shape of the *Dotted Record of Canton*. This record is contained in the *Li-tai san-pao chi* written by Fei Chang-fang in 597. This source, as discussed by W. Pachow in "A Study of the Dotted Record," mentions that, according to the famous Buddhist Master Samghabhadra:

> there is a tradition which had been handed down from teacher to teacher for generations, viz., after the passing away of the Buddha, Upali collected the Vinaya and observed the Pavarana on the 15th of the 7th Moon of the same year. Having offered flowers and incense to the Vinaya on that occasion, he marked a dot [on a record] and placed it close to the Vinaya text. Thereafter this was repeated every year. . . . In this manner the teachers in turn handed it down to the present master of *Tripitaka*. . . . Having observed the *Pavarana* and offered flowers and incense to the *Pavarana* at midnight (on the

15th) of the 7th Moon, in the 7th year of Yung-ming [AD 489], he added a dot [to the record]) as a traditional practice. The total amounted to 975 dots in that year. A dot is counted as a year. (1965: 343)

On the basis of the figures supplied in this record, we get 489 CE minus 975 years equals 486 BCE as the year of the Mahaparinirvana. But Pachow noted that three extra dots had been inadvertently added. The actual number of dots in the year 489 CE should have been 972 and not 975. In that case, he pointed out, the actual date of the *Mahaparinirvana* should be 489 CE minus 972 equals 483 BCE.

As the two independent sources of information, one from Sri Lanka (as mentioned in the *Dipavamsa* and the *Mahavamsa*) and the other from China (the *Dotted Record of Canton*), provided substantially the same information, about 483 BCE was accepted as the correct date of the *Mahaparinirvana* of the Buddha by Buddhist scholars.

Other Support for Long Chronology

The long chronology has also been supported on the basis of the so-called agreement of this chronology with the Jaina chronology as well as the *Puranas*. The Pali Canon points out clearly that the Buddha and the Mahavira were contemporaries. Since an apparently independent, although late, Jaina tradition states that the death of the Mahavira took place 155 years before the accession of Chandragupta, and since the accession of Chandragupta can be dated to about 317 BCE, Mahavira Jaina's death may be put in the year 317 plus 155 equals 472 BCE. But here the main difficulty is that the same Pali source, the *Digha Nikaya*, places Mahavira Jaina's death before that of the Buddha. Two separate answers have been provided for this contradiction. One, as pointed out by Hermann Jacobi (1879) in his introduction to *Kalpasūtra of Bhadrabāhu*, is that the Buddhist texts were confused by there being two places called Pava and were probably also confused by the relative dating. The second is that the southern Buddhists, as proposed by A. L. Basham (1951) in *History and Doctrine of the Ājivikās*, knew very little about other sects, and it was the Ajivika leader Makkhali Gosala who had died before the Buddha and not Mahavira Jaina.

The long chronology has also found strong support in the information available in the edicts of King Ashoka. For instance, the Minor Rock Edict (MRE) I of Ashoka, which refers to the date 256. As noted by G. Bühler (1877), this figure has been interpreted by these scholars to mean a time span of 256 years between the installation of MRE I and the *Mahaparinirvana* of the Buddha. A. K. Narain has discussed in detail the implications of the number 256 and has vigorously proposed that it is clinching evidence for proving that the Buddha's *Mahaparinirvana* took place about in the year 483 BCE. He has translated the relevant portion of the edict as follows: "This proclamation [was made] having given [that is, allowed or having past] two hundred and fifty-six [years] to

---

Mahavira Jaina

Also known as Vardhamana, Mahavira (599–527 BCE) was the central figure of Jainism, which grew and developed in parallel to Buddhism. Mahavira was said to be the 24th of the Tirthankaras, perfectly enlightened beings. The first 23 Tirthankaras may have been mythical or at least exaggerated by legend—they're described as hundreds of feet tall, living for thousands of years.

Mahavira taught that karma accumulates on the soul in response to good and evil deeds, and the soul responds by seeking the temporary pleasures of the material world. He taught the necessity of five vows in order to lead a life of right conduct: nonviolence, truthfulness, abstinence from theft, chastity, and detachment from the material world.

---

elapse [after] the ascension of the body of our Buddha" (Narain 2003). The date of issue of this edict is hard to fix, and Narain feels that it must have been issued toward the end of Ashoka's reign, that is, in the 37th year. This means the edict was issued in the year 228 (265 minus 37) BCE. The upshot of this is that the *Mahaparinirvana* of the Buddha took place about in the year 484 (228 plus 256) BCE. As months and days are not mentioned, about 484 can be recalibrated to about 483 BCE. A date for the Buddha calculated in this manner has its own merits. As pointed out by Narain, it "is independent of the so-called two Buddhist traditions as well as that of the *Dotted Record*, the amended version of which, incidentally stands not only substantiated now but also freed from its dubious association with the *later* Theravada tradition" (Narain 2003).

Archaeological evidence also appears to support the long chronology. The Pali Canon gives clear evidence of Buddhism being an urban religion and the Buddha having preached in urban centers. A large number of the urban centers mentioned in the Pali Canon

Dhamekh Stupa at Sarnath in Uttar Pradesh, India, is a stone and brick structure built during the reign of Emperor Asoka in the third century BCE. The stupa, constructed on the site where the Buddha (Siddhartha Gautama) preached his first sermon, is decorated with intricate floral patterns. (Philip Baird (http://www.anthoarcheart.org)

that have been identified have provided evidence of northern black polished ware (NBPW). Though the precise dates of the origin and spread of NBPW in the Ganges Valley are not without dispute, there is a general consensus that it had become fairly widespread by about 500 BCE. It would be fair to say that the urbanization of the Ganges Valley, also sometimes called the Second Urbanization, originated in the sixth century BCE. Buddhism can also be traced back to at least 550 BCE. Archaeological records relating to the excavation records of some of the urban centers in the Ganges Valley also tend to support this.

Here, it may not be out of place to look at the archaeological records of Kaushambi, the oldest city of the Second Urbanization. This city appears to have been established by at least the end of the Vedic period, though its excavator, G. R. Sharma, places it as early as 1000 BCE. According to the Buddhist canonical text, the *Digha Nikaya*, Kaushambi was a well-known capital city of the Vatsas/Vamsas and was one of the six major cities (*mahanagaras*) of India at the Buddha's time. Major trade routes of the time passed through this city. Kaushambi was perhaps one of the most important cities politically, religiously, and economically at the time of the Buddha. The Mathura sculpture from the Ghoshita Monastery of a *Chakravarti* Buddha of year two of Kanishka I, installed according to the inscription at the promenade of Gautama Buddha, is the oldest Buddhist relic from Kosam. With the help of an inscribed stone slab the monastery was identified with the well-known Ghoshita Monastery. The excavator, Sharma (1960), places the first phase of its construction in about 600 BCE.

Shravasti was perhaps the most important city for the Buddha, considering he delivered the largest number of his sermons in this town and spent most of his Rainy Retreats (*vassavasa*) here. Archaeologists, like K. K. Sinha (1967), who have either excavated it or who, like H. Härtel (1991), have studied the data available on this city, have pointed out that the origins of this city go at least as far back as the sixth century BCE.

Even the earliest portions of the Pali Canon presuppose the existence of a developed currency, and such a currency involving large transactions of gold and silver coins must have taken time to develop. Although it has been debated whether the earliest coins can be dated, as P. L. Gupta discusses, "coins . . . were current prior to the fifth century BC" (Gupta 1969: 11). Though no evidence of coinage can be found in later Vedic texts, measures of precious metals may have been used as payment. Discovery of 3,000 cowrie shells from the NBPW levels at Masaon-Dih throws interesting light on the use of currency before the introduction of coins. Without entering into discussion on the numismatic evidence, it may be reasonable to assume that coins made their beginning in India during the sixth century BCE. Thus the evidence of the existence of coinage also seems to support the long chronology.

Though the stratigraphical sequence of the cultures of the Ganga Valley is now well established, the absolute chronology still remains debatable. In "Radiocarbon

Dates from South Asia," G. L. Possehl (1987) notes that now quite a few radiocarbon dates from various sites are available. Though normally they should suffice for establishing the chronology of various cultures, the erratic nature of many dates (even after calibration) has divided archaeologists nearly as much as have the two traditions for the date of the Buddha. While dealing with C14 dates, we also have to bear in mind several associated problems, especially, as D. H. Thomas (1978) notes, that they are not precise statements of the age of samples but estimates of probability. It is unlikely that we will get uniform dates for the beginning and end of a culture from all parts of its geographic area. The Buddhist order depended on the existence of a strong economic base. The monks were supposed to spend the Rainy Retreat in fixed locations, and this would have been easiest near large urban settlements. The large cities were no longer mere administrative centers and sovereign residences. They had also become the nerve centers of economy and commerce. Uncertain and unsatisfactory as archaeological data still are in this context, they appear to lean toward supporting an early rather than late date for the *Mahaparinirvana* of the Buddha. There thus is at least a good case that can be made for the Buddha having lived in the sixth century BCE.

Ceylon-India Chronological Connections

In the chronological system, the succession of the great teachers from Upali onward plays an important role. There is a continuous synchronological connection between the history of Ceylon and that of India. The *Dipavamsa* and the *Mahavamsa* talk of five patriarchs (*acharyas*) who transmitted the *Vinaya* from the time of the Buddha's death until the days of Ashoka. These five elders were Upali, Dasaka, Sonaka, Siggava, and Moggaliputta Tissa. The *Dipavamsa* mentions: "Seventy-four of Upali, sixty-four of Dasaka, sixty-six of Thera Sonaka, seventy-six of Siggava, eighty of Moggaliputta: this is the *Upasampadā* of them all" (1958).

Though this verse mentions the years of *Upasampada*, in reality these are the ages at which these patriarchs died. This fact is borne out by the verses preceding as well as following this verse. For instance, the *Dipavamsa* mentions that Upali attained nirvana at the age of 74. Thus, the number 74 mentioned in the verse is the age at which Upali died and not the year or period of *Upasampada*. The same should be taken to be the case regarding the other numbers mentioned in connection with the other elders.

The other verse of the *Dipavamsa* says, "Learned Upali was all the years chief of the Vinaya, Elder Dasaka fifty, Sonaka forty-four, Siggava fifty-fifth year, the [elder] called Moggaliputta sixty-eight" (1958). This verse clearly implies the number of years for which the five elders were the custodians of the *Vinaya*.

It appears that Upali joined the order at quite a mature age. He was born in the family of a barber, later took up service with the Sakyan princes, and joined

the order along with them. Even during the lifetime of the Buddha, monks considered it a great privilege to learn the *Vinaya* under him. He specialized in the study of the *Vinaya* and won the foremost place among the *Vinayadharas*. According to the *Dipavamsa*, he was renowned for having reached the pinnacle of the *Vinaya*, and it was in this capacity that Kassapa entrusted him with compiling the *Vinaya Pitaka* at the First Buddhist Council that took place at Rajagriha. We are further told in the *Dipavamsa* that when 16 years had elapsed after the death of the Buddha, Upali was 60 years old. This means that he was 44 (60 minus 16) years old when the Buddha died, that is, when he became the *Vinaya* custodian. But as mentioned above, he actually lived to be 74. Thus, Upali was the custodian of the *Vinaya* for 30 (74 minus 44) years. This is also supported by a direct statement in the *Dipavamsa* that Upali was the custodian of the *Vinaya* for 30 years.

Dasaka was a learned brahmana from Vesali. After meeting and holding a discussion with Upali, Dasaka entered the order to study the doctrine. He appears to have been fairly mature in years when he joined the order. He learned the whole of the Vinaya and became an *arahant*. As per the *Dipavamsa*, he was the custodian of the Vinaya for a period of 50 years and was followed by Sonaka, the son of a caravan leader from Kasi, who had joined the order at the age of 15 at Rajagriha. He saw Dasaka Thera, and, very pleased with him, entered the order after fasting for three meals until his parents would give their consent. He soon became an *arahant* and leader of 1,000 monks. Sonaka kept the *Vinaya* for 44 years.

Siggava, the son of a minister from Pataliputra, joined the order at the age of 18 along with his friend Chandavriji. As pointed out in the *Dipavamsa*, Siggava was the custodian of the *Vinaya* for 54 years (having died during the 55th year of custodianship). Siggava died when 14 years of the reign of Chandragupta had elapsed. As King Chandragupta Maurya had begun his reign in about 321 BCE, Siggava's death took place in about the year 321 minus 14, which is about 307 BCE.

Conclusion

The above-stated information based on various archaeological and literary sources may be summed up as follows.

The lifespan of the Buddha is arrived at by adding together two numbers, one being the date of the accession of Ashoka to the throne, the second being the length of the interval between that date and the date of the death of the Buddha. Upali, Dasaka, Sonaka, and Siggava kept the *Vinaya* for 30, 50, 44, and 54 years, respectively. The death of Siggava took place in about the year 307 BCE. Between about 307 BCE and the death of the Buddha, 178 years had elapsed. As the custodianship of these four patriarchs is mentioned only in years, and months and days are not mentioned, an error of a couple of years is possible. Considering this, it may not be out of order to adjust the figure of 178 to 176. This would mean that the Buddha's death may approximately be placed in about the year 483 BCE. This agrees with the date

calculated on the basis of the popular number 218 as well as the *Dotted Record of Canton*. Thus the year 483 should be accepted as the year in which the death of the Buddha took place. However, this date should only be taken as a close approximation to the real date rather than an exact date for the reasons specified above.

References and Further Reading

Bareau, A. "La date du nirvāna." *Journal Asiatique* 241 (1953): 27–62.

Basham, A. L. *History and Doctrine of the Ājivikās*. London: Luzac, 1951.

Bechert, H. "The Date of the Buddha Reconsidered." *Indologica Taurinensia* 10 (1982): 29–36.

Bechert, H., ed. *The Dating of the Historical Buddha/Die Datierung des historischen Buddha*, 2 vols. Göttingen, Germany: Abhandlungen der Akademie der Wissenschaften, 1991–92.

Bechert, H., ed. *When Did the Buddha Live?* Delhi: Sri Satguru Publications, 1995.

Bühler, G. "Three New Edicts of Aśoka." *Indian Antiquary* 6 (1877): 149–60.

Dipavamsa, The. Edited and translated, with introduction, by B. C. Law as "The Chronicle of the Island of Ceylon or the Dīpavamsa." *Ceylon Historical Journal* 7 (1958): 1–266.

Eggermont, P. H. L. *The Chronology of the Reign of Asoka Moriya*. Leiden: E. J. Brill, 1956.

Eggermont, P. H. L. "New Notes on Aśoka and His Successors II." *Persica* 4 (1969): 97.

Geiger, W., ed. *The Mahāvamsa*. London: Pali Text Society, 1912. Translated by W. Geiger and M. H. Bode as *The Mahāvamsa or the Great Chronicle of Ceylon*. London: PTS, 1912.

Gombrich, R. "Dating the Buddha: A Red Herring Revealed." In *The Dating of the Historical Buddha/Die Datierung des historischen Buddha* (Vol. 1, pp. 238–59). Edited by Heinz Bechert. Göttingen, Germany: Abhandlungen der Akademie der Wissenschaften, 1991.

Gupta, P. L. *Coins*. New Delhi: National Book Trust, 1969.

Härtel, H. "Archaeological Research on Ancient Buddhist Sites." In *The Dating of the Historical Buddha/Die Datierung des historischen Buddha* (Vol. 1, pp. 61–89). Edited by Heinz Bechert. Göttingen, Germany: Abhandlungen der Akademie der Wissenschaften, 1991.

Jacobi, H., ed. "Introduction." In *Kalpasūtra of Bhadrabāhu*. Leipzig, 1879.

Mendis, G. C. "The Chronology of the Early Pāli Chronicle of Ceylon." *University of Ceylon Review* 5, 1 (1947): 39–54.

Müller, Max F. "The True Date of Buddha's Death." *The Academy* (March 1884): 153.

Narain, A. K. "An Independent and Definitive Evidence on the Date of the Historical Buddha." *Buddhist Studies Review* 6 (1994), 43–58.

Narain, A. K. *The Date of the Historical Śākyamuni Buddha*. Delhi: B.R. Publishing, 2003.

Pachow, W. "A Study of the Dotted Record." *Journal of the American Oriental Society* 83, 3 (1965): 342–45.

Possehl, G. L. "Radiocarbon Dates from South Asia," data list circulated by the author in September 1987.

Rhys Davids, T. W. "The New Aśoka Inscriptions." *The Academy* (July 1877): 37.

Rhys Davids, T. W., and J. E. Carpenter, eds. *The Dīgha Nikāya*, 3 vols. London: Pali Text Society, 1890–1911. Translated by T. W. and C. A. F. Rhys Davids as *The Dialogues of the Buddha*, 3 vols. London: Pali Text Society, 1899, 1910, and 1957.

Sarao, K. T. S. "Urban Centres and Urbanization as Reflected in the Pāli Vinaya and Sutta Pitakas." Ph.D. thesis submitted to the University of Cambridge, 1989.

Sharma. G. R. *The Excavations at Kausambi: 1957–59*. Allahabad, India: University of Allahabad, 1960.

Sinha, K. K. *Excavations at Sravasti: 1959*. Varanasi, India: Banaras Hindu University, 1967.

Stein, O. "The Coronation of Candragupta Maurya." *Archiv Orientalni* 1 (1932): 368.

Takakusu, J. "Pāli Elements in Chinese Buddhism: A Translation of Buddhaghosa's *Samantapāsādikā*, a Commentary on the *Vinaya*, Found in the Chinese *Tripitaka*." *Journal of the Royal Asiatic Society* (1896): 436.

Thomas, D. H. "The Awful Truth about Statistics in Archaeology." *American Antiquity* 43 (1978): 232.

Thomas, E. J. "Theravādin and Sarvāstivādin Dates of the Nirvāna." In *B.C. Law Volume* (Vol. 2, pp. 18–22). Edited D. R. Bhandarkar et al. Delhi: Poona, 1946.

Wickremasinghe, D. M. Z. "Kiribat-Vehera Pillar Inscription." *Epigraphia Zeylanica* 1 (1904–12): 153–61.

Winternitz, M. *A History of Indian Literature*, Vol. 2. Calcutta: University of Calcutta, 1933.

Yamazaki, G. "The Importance of the Dotted Record." In *The Date of the Historical Śākyamuni Buddha* (pp. 147–50). Edited by A. K. Narain. Delhi: B. R. Publishing, 2003.

CON

Though there is general agreement that the Buddha lived for 80 years, precisely when he lived is hard to pinpoint for three main reasons. First, the different texts that give information on the birth and death dates of the Buddha not only contradict each other, but they are also self-contradictory in most cases. Second, because of their religious character, most of these texts do not mention numbers in the sense in which they are understood in science. In other words, in most cases the numbers can only be used as approximations. In many cases the numbers are also given as rounded-off numbers and hence cannot be used for precise calculations. Invariably the Buddhist texts appear to exaggerate numbers, and in all Indian religions, there is always a tendency to claim antiquity for a religious leader. Moreover, as the textual sources mention, spans of time are given in years only, not months and days; the figures are not as precise as one would want them to be. Third, no useful date for calculating the lifetime of the Buddha is available in the history of ancient India before the arrival of Alexander the Great. Because of these factors, it is doubtful at best that the traditional dates for Buddha's life, 563–483 BCE, are accurate.

The calculation of the dates of the Buddha is inextricably linked to the dates of the Mauryan kings, Chandragupta and Ashoka. Therefore, these two dates need to be determined before any work can be done on the dates of the Buddha. It is more or less certain that Chandragupta started to rule in about 317 BCE, although some scholars have put it a little earlier. According to N. K. Bhattasali in "Mauryan Chronology and Connected Problems," "The murder of Poros by Endamos, and his retirement from India in 317 BCE are significant indications. The breaking out of the Indian revolt headed by Chandragupta does not appear to be possible before this date" (1932: 283), and, therefore, according to O. Stein in "The Coronation of Candragupta Maurya," it "is impossible to reckon with an acknowledged dominion of Candagutta before 317 BCE" (1932: 368). On the basis of the names of various Greek kings mentioned in the Thirteenth Rock Edict, the date of Ashoka's accession may be put in about 268 BCE and the consecration, which took place in the fourth year of his reign (i.e., after three years), in about 265 BCE.

Long and Short Chronologies

The sources used for the study of the dates of the Buddha may broadly be divided into two categories, depending on whether they support the so-called long chronology or the short chronology. These chronologies are based mainly on the southern and northern Buddhist legends, respectively.

The southern Buddhist legends contained in the *Dipavamsa* and the *Mahavamsa* place the consecration of Ashoka 218 years after the *Mahaparinirvana*

(the Great Decease) of the Buddha. The date of the Buddha's death would therefore have taken place in about 483 (265 plus 218) BCE. The short chronology is based on the testimony of the Indian sources (*Vinaya Pitaka*) and their Chinese and Tibetan translations. In all the recensions of the *Vinaya Pitaka*, it is pointed out that the Buddha died 100 or 110 years before the consecration of Ashoka. In other words, the *Mahaparinirvana* should be dated in about 365 (265 plus 100) BCE or 375 (265 plus 110) BCE.

One possible important reason for the popularity of the long chronology is that, instead of the suspicious number of 100 in the short chronology, the long chronology has the exact number of 218. But this does not necessarily mean that 218 is a true number just because it does

A stone relief of the Buddha's footprints on a pillar of the stupa commissoned by the emperor Asoka in the third century BCE at Sanchi in present-day India. (Adam Woolfitt/Corbis)

not appear to be rounded off. It is also important to note that the weakness of the long chronology is that the *Dipavamsa* (fourth to fifth centuries CE) in which it is found was written two or three centuries later than the sources in which the short chronology first appears. The longer the interval between the time of the events and the time they were recorded, the greater the possibility of an objective error. The number 218 may not be acceptable on various other grounds too. For instance, it may have been inflated through additions to an originally much smaller number so that credence could be given to various personalities as well as events. As a matter of fact, the long chronology appears to have been developed in an attempt to adjust the traditional short chronology to the particular needs of the Sri Lankan historiography. Thus, as pointed out by E. J. Thomas (1946) in "Theravadin and Sarvastivadin Dates of the Nirvana," the relevant passages in the *Dipavamsa* actually point to the existence of the original short chronology, which failed to be assimilated into the long chronology of the final version of the *Dipavamsa*.

W. Geiger's (1912) discussion of the chronology of the Buddha appears to have been extremely influential in the acceptance of the long chronology over the short chronology. Other scholars like André Bareau (1953, 1995) and P. H. L. Eggermont (1956, 1969) followed suit, and thus the long chronology became

the basis for the date of the Buddha. However, the biggest justification for the long chronology came in the shape of the *Dotted Record* of Canton (Guangzhou), contained in the *Li-tai san-pao chi*, written by Fei Chang-fang in 597 CE. According to this tradition of putting one dot on the Vinaya record every year resulted in 975 dots in the year 489 CE, as detailed by W. Pachow (1965) in "A Study of the Dotted Record." But Pachow believed that three extra dots might have been inadvertently added, and that the actual number of dots should have been 972 and not 975. On this basis, 483 BCE (i.e., 489 CE minus 972 years) was calculated as year of the *Mahaparinirvana*.

But this tradition from the Chinese sources is apparently not independent in origin. It has been maintained, for instance by A. Bareau and J. Takakusu (1896), that this tradition initially originated in Sri Lanka and hence cannot be used reliably. It appears thus that the dot is a later invention to dignify the *Vinaya*. Moreover, the very way in which it was preserved, handed down from generation to generation and carried from one country to another, appears rather mysterious and suspicious. We cannot but express doubts concerning its authenticity. Most important, the Sri Lankan chronicles and the *Samantapasadika* speak of the transmission of the *Vinaya* by the teachers initiated by Upali, but in them we do not come across any reference, whatsoever, to the practice of adding dots to a record every year after the Rainy Retreat (*vassavasa*). Such being the case, it is difficult, according to Pachow, to believe that the dotted record was initiated by Upali and handed down in succession by the *Vinaya* teachers. Moreover, if there was really a record initiated by Upali, when Mahinda, the sixth teacher of the *Vinaya* succession, came to Sri Lanka, he should have brought it with him and continued to add dots each year throughout his life. If so, such a record would have been safely preserved in Sri Lanka as a sacred object like the Bo-tree, or the Tooth Relic. But this was not known to writers of either the Pali or the Sri Lankan texts, nor was it noted in the *Travels of Fa-hsien*, when Fa-hsien (Faxian) visited Sri Lanka in the beginning of the fifth century. Thus one may pose the question whether Mahinda really brought such a thing to Sri Lanka. In case such a thing did not exist in Sri Lanka, then one may ask how it came to China, and from where. In any case, as no written record of the *Vinaya* existed until the time of Dutthagamani in the first century BCE, it is difficult to accept the authenticity of this tradition. Moreover, as Max Müller notes in "The True Date of Buddha's Death," "the process of adding one dot at the end of every year during 975 years is extremely precarious" (1884: 153).

The long chronology has also been supported on the basis of the so-called agreement of this chronology with the Jaina chronology as well as the *Puranas*. But the *Puranas* show so many disagreements among themselves that they are not really reliable for calculating the dates of the Buddha. The most important reason for not using the Jaina chronology for dating the Buddha, according to Bareau (1995), is that the Jaina chronology itself depends on certain Buddhist traditions,

notably the Sri Lankan tradition. Thus, despite the fact that the two teachers were contemporaries, it is difficult to accept the Jaina chronology for its inherent snags. According to M. Winternitz (1933) in *A History of Indian Literature*, the tradition of the long chronology cannot be traced with confidence beyond the middle of the 11th century. Some scholars have supported the long chronology on the basis of three Ashokan edicts of Sahasaram, Rupanath, and Bairat, which refer to the year 256. This figure has been interpreted by such scholars as G. Bühler (1877) in "Three New Edicts of Ashoka" to mean a time span of 256 years between the installation of these inscriptions and the *Mahaparinirvana*. An attempt has also been made by scholars to present a date akin to the short chronology on the basis of these inscriptions. For example, T. W. Rhys Davids (1877: 37) provided "426 BCE, or perhaps a few years later" as the date of the *Mahaparinirvana* by pointing out that the number 256 represents the time span between the installation of these inscriptions and the Buddha's abandonment of his home. However, some scholars, like Hermann Oldenberg (1881), have pointed out that unnecessarily too much has been made of this figure not only because the inscriptions contain no word for years but also because they do not refer to the Buddha, but to 256 "beings."

The theory of 100 years is widespread throughout the world. Geiger notes that the Tibetan sources place the reign of Ashoka 100 to 160 years after the Buddha's death. Taranatha says that the Tibetan *Vinaya* gives 110 AB (after Buddha) as one of the dates for Ashoka. Similarly the Chinese *Tripitaka* gives 116, 118, and 130 AB as the dates for the consecration of Ashoka. In Vasumitra's account, Ashoka is also placed about 100 year after the death of the Buddha. According to Hsuan-tsang (also spelled as Xuanzang), as noted by S. Beal (1906) in *Si-yu-ki: Buddhist Records of the Western World*, at the time of his death, the Buddha had said "A hundred years hence there shall be a King Ashoka." Still at another place, Hsuan-tsang points out "the different schools calculate variously from the death of the Buddha. Some say it is 1,200 years and more since then. Others say, 1,300 or more. Others say, 1,500 or more. Others say that 900 years have passed, but not 1,000 since the nirvana."

The various dates here recorded would correspond with 552, 652, 852, and a date between 252 and 352 BCE. By the last date, Hsuan-tsang probably means to place the death of the Buddha a hundred years before Ashoka.

Case for a Later Date

Two important reasons, however, appear to favor a later date for the Buddha. They are the archaeological considerations and the lists of the patriarchs (*acariyaparampara*). The archaeological records in the Ganges Valley show (perhaps with the exception of Kaushambi) that even by about 450 BCE, the new urban settlements were indeed not those cities we might expect after reading early Buddhist literature. Extensive use of baked bricks for construction, a well-

developed sanitation system, and so forth are not found in the excavations until later times. In early Buddhist literature, the existence of prosperous and fully developed urban centers is taken for granted. Though the roots of the Ganges urbanization may be traced back to about 500 BCE or so, the archaeological records clearly suggest that the sort of urban centers that are talked about in the earliest Buddhist texts could not have come into existence before the end of the fifth century BCE. Critics of this argument may say that such references are later interpolations or that certain portions of the canon are altogether late compositions. But such criticism appears superficial, because the whole material milieu reflected in early Buddhist literature is urban. Wherever we may look, Pali *Tipitaka* reflects a city culture and a faith laden with munificence by the city folks that included kings, their ministers, and business magnates.

As many as 173 urban centers (some undeniably being mythical or late) are mentioned in the first two *pitakas* and are evenly spread out in these texts. Here an argument may be made that perhaps the whole of Buddhist literature was grafted onto various urban settlements for prestige or other reasons, because terms associated with village (*gama*), such as *gamadhamma* (vile conduct) and *gamakatha* (village-talk, included in the list of foolish talks), are frowned on in Buddhist literature. But it is impossible to accept such an argument. It is not only the urban settlements but so much else that goes into making an urban civilization that is reflected everywhere in early Buddhist literature. Long-distance trade, a money economy, financial transactions, interest, usury, mortgages, the developed state and its paraphernalia, prostitution, and many other characteristics clearly point to the existence of a fully grown urbanization in Buddhist literature. There is so much urbane that is part and parcel of the life and activities of Gautama Buddha, it would be hard to imagine him living in a preurban society.

A part of the *Mahaparinibbana Sutttanta* of the *Digha Nikaya*, which mentions six *mahanagaras* (cosmopolitan cities), is dated by Winternitz as forming part of the earliest Buddhist literature. These *mahanagaras* were Champa, Rajagriha, Shravasti, Kaushambi, Saketa, and Benares. A look at the scanty evidence so far provided by the excavators of these cities clearly tempts agreement with the short chronology. If we are to accept the existence of these six settlements as *mahanagaras*, then that can be visualized perhaps by the end of the fifth century BCE at the earliest. The archaeological data available from the Ganges Valley show that even by about 500 BCE, the new urban settlements were indeed not those cities that may be expected after reading the early Buddhist literature.

Though scholars disagree as to when coins came into existence in India, it is reasonable to say they were introduced in India during the fifth century BCE. Even the earliest portions of the Pali Canon presuppose the existence of a developed currency, and such a currency involving large transactions of gold

and silver coins must have taken time to develop. The Buddhist Sangha depended on the existence of a strong economic base. The monks were supposed to spend the Rainy Retreat in fixed locations, and this would have been easiest near large urban settlements. The large cities were no longer mere administrative centers and sovereign residences. They had also become the nerve centers of economy and commerce. Uncertain and unsatisfactory as archaeological data still are in this context, they appear to lean toward supporting a later rather than earlier date for the *Mahaparinirvana* of the Buddha. In other words, there is at least a good case that can be made for the age of the Buddha being about a century later than generally accepted.

As pointed out above, extensive use of bricks for construction works, including fortifications, well-developed sanitation, palatial buildings, a fully developed state system and its paraphernalia, an extensive interregional commercial network with powerful and influential business magnates, a well-developed currency and other financial institutions like usury, mortgage, and so forth, is well reflected throughout the Pali *Tipitaka*. The material milieu reflected in the early Buddhist literature is overwhelmingly urban. A collective analysis of the data available on the six *mahanagaras*, mentioned in the earliest portions of the Pali literature, shows that urban centers of this magnitude could not have existed before the end of the fifth century BCE. As compared to the later Vedic texts and their socioeconomic context, the early Buddhist texts depict a prosperous urban life, a flourishing interregional trade dominated by a new class of influential and powerful merchants, and the emergence of Magadha as the most powerful early state among a large number contesting *mahajanapadas* in the Ganges Valley.

The existence of fortifications around the various urban centers and their relationship with the Buddha's time constitute yet another problem difficult to resolve for an early date. The archaeological evidence does not support the fortification of any of the early Ganges cities, with the possible exception of Kosambï, even in the fifth century BCE, whereas fortified towns are frequently mentioned in the early Buddhist texts.

Political power, centered in the urban centers, and riches were accumulated in these cities. The emergence of these strong *mahajanapadas*, which is identifiable mainly in the early Buddhist literature, therefore would have to be dated in the fifth century BCE rather than in the sixth century BCE, as has been the custom in dating them until now. Furthermore, such an interpretation would provide the needed time for a gradual evolution of the urban settlements and their surrounding kingdoms. The same would be true with regard to the development of interregional trade and the rise of an urban merchant class. The latter, in particular, may have needed much more time than we have conceded to them in view of the early date of the Buddha and of the early Buddhist literature, which depicts an already flourishing merchant culture. Such a late date of the rise of urban centers, a merchant class, and its flourishing interregional trade may help to explain the lateness of the punch-marked coins.

Some scholars believe that a consideration of the probable distance between the Buddha and Ashoka in terms of doctrinal development of Buddhism, as L. Schmithousen writes, "would seem to render a somewhat later date more probable" (1992: 143). A study of Buddhist poetry also tends to show, as S. Lienhard writes, that the corrected long chronology "definitely seems to lie too far back in time" (1991: 196). "It would seem to be easily compatible with the assumption that Buddhism had not yet produced distinctive monuments and institutions, and that, instead, it was still rather young and not yet fully visible when Megasthenes visited Pataliputra around 300 B.C." (Halbfass 1995: 205). P. H. L. Eggermont, in his "New Notes on Ashoka and His Successors II," also feels that "Buddhism was still young at Ashoka's time" (1956).

In the chronological system upon which the *Dipavamsa* and the *Mahavamsa* are based, the succession of the great teachers from Upali down to Mahinda played an important part. This *acariyaparampara* is of interest because in it there is a continuous synchronological connection between the histories of Sri Lanka and India. Here the system appears to have been carried out in detail and completed. As is clear in the accounts of the *Dipavamsa* and the *Mahavamsa* there was a teacher–pupil relationship between them, and this continuity is of vital importance. The lists of *acariyas* that occur in the *Vinaya*, Sri Lankan chronicles, and elsewhere as *Vinayadharas* are more reliable and useful than any other form of information to determine the date of the Buddha. As most of the research was conducted in the light of number 218, it was given that the number of elders as the *Vinayapamokkhas* for the period between the Buddha and Ashoka caused a problem. There were not enough elders. Thus it was pointed out that to bridge the gap of 218 years, each of the elders had to be assigned too lengthy a period of time as guardian of the *Vinaya*. The statement in both the *Dipavamsa* and the *Mahavamsa* that the eight elders who considered the Ten Extravagances in the Second Council had all seen the Buddha was also seen as creating difficulties. These so-called *contradictions*, however, were regarded as faulty records on the part of the Theravadins. More weight was given to the chronology of the kings, even though this, too, posed difficulties. All these problems had come up because the number 218 was thought to be supreme.

In our calculation of the date of the Buddha based on the lists of patriarchs, we have used the beginning of the reign of Chandragupta as the base year as opposed to the year of Ashoka's coronation. This shortens the gap between the date of the Buddha and the base year, thus reducing the margin of error.

Patriarchs

According to E. Frauwallner (1956) in *The Earliest Vinaya and the Beginnings of Buddhist Literature*, northern sources (the *Divyavadana*, the *Ashokavadana*, and so forth) point out three generations of patriarchs, that is, Mahakassapa/

Ananda, Sanavasa, and Upagupta, dating from the Buddha's death to the time of Ashoka (excluding Madhyantika, whose name appears to have been inserted by the legend-teller monks). Sanavasa was a merchant of Magadha at the Buddha's time, who after the Buddha's death became a monk under Ananda's guidance, moved to Madhura (Mathura) later on, and introduced Upagupta into monkhood. Sanavasa must be Sambhuta-Sanavasi of Madhura/Ahoganga, who took part in the Second Council. As many different sects agree, it appears Sanavasa's participation in the Second Council is quite probable. Upagupta is said to have been a temporary advisory monk of Ashoka.

The southern sources relate that five patriarchs transmitted the *Vinaya* from the time of the Buddha's death until the days of Ashoka. These five elders were Upali, Dasaka, Sonaka, Siggava, and Moggaliputta Tissa. We are told in the *Dipavamsa* that when 16 years had elapsed after the death of the Buddha, Upali was 60 years old. This means he was 44 (60 minus 16) years old when the Buddha died, that is, when he became the *Vinayapamokkha*. As Upali lived to be 74, he was the custodian of the *Vinaya* for 30 (74 minus 44) years. This is also supported by a direct statement in the *Dipavamsa* that Upali guarded the *Vinaya* for 30 years.

When Upali died, Udaya had completed 6 years of his 16-year reign. This means during the last 10 (16 minus 6) years of Udaya's reign, Dasaka was the custodian of the *Vinaya*. But Dasaka died when 8 years of the 10-year reign of Susunaga had elapsed. As Anuruddhaka/Munda ruled for 8 years between Udaya and Susunaga, Dasaka appears to have been the custodian for a total of 26 years (10 plus 8 plus 8).

Susunaga ruled for 10 years and Dasaka died 8 years after the end of Susunaga's reign. After the death of Susunaga, the Ten Brothers reigned for 22 years, and Sonaka died when 6 years of their reign were over. This means Sonaka kept the *Vinaya* during the last 2 years of the reign of Susunaga and first 6 years of the reign of the Ten Brothers, making it 6 (2 plus 6) years.

Siggava was the custodian during the remaining 16 (22 minus 6) years of the reign of the Ten Brothers. Siggava died when 14 years of the reign of Chandragupta had elapsed. In other words, Siggava was the custodian for a total period of 30 (16 plus 14) years.

Chandragupta does not appear to have succeeded the Ten Brothers, who began their reign not at Pataliputra but elsewhere, because the *Dipavamsa* and the *Mahavamsa* tell us that Susunaga had a son called Kalashoka who held power at Pataliputra for a period of 28 years. It appears after his governorship for 10 years during Susunaga's reign, Kalasoka reigned for 18 years (28 minus 10) as a king at Pataliputra, and the Ten Brothers continued to rule from the same place as Susunaga after the possible division of the kingdom. In other words, it appears that Chandragupta succeeded Kalashoka at Pataliputra and the Ten Brothers (possibly the Nandas) at Rajagriha. The *Dipavamsa* also tells us that Siggava was 64 years old when Chandragupta had completed two years of

---

Chandragupta Maurya

The founder of the Maurya Empire, Chandragupta was the first to unite the lands and peoples of the Indian subcontinent. He first rose to fame by reconquering the Indian lands his contemporary Alexander the Great had taken over, and soon conquered the Nanda Empire and expanded to the east. The Maurya Empire ruled by his dynasty was the most powerful period of ancient India, and the greatest to rule the subcontinent until India became a British subject for a (relatively) brief time in the modern era. It was a time of religious awakening for the subcontinent, not only through the birth of Buddhism but through Chandragupta's own conversion to Jainism. In his last days, the emperor voluntarily resigned from the throne and finished out his life as a Jain ascetic, fasting in a cave.

---

his reign. Chandragupta's reign began in about 317 BCE. This means that in about 315 (317 minus 2) BCE, Siggava was 64 years old. But as Siggava died at the age of 76, that means he lived for another 12 years after 315 BCE. This would put the death of Siggava in about 303 BCE. This statement is also supported by another reference in the *Dipavamsa* where we are told that Siggava died 14 years after the beginning of the reign of Chandragupta, that is, about 303 BCE.

Conclusion

The upshot of the calculations made above is that the death of Siggava took place in about 303 BCE. Sonaka died 30 years before Siggava. Dasaka died 8 years before Sonaka. Upali died 26 years before Dasaka. The Buddha died 30 years before Upali. In other words, between about 303 BCE and the death of the Buddha, 94 years had elapsed. This would mean that the Buddha died in about 397 BCE.

It must finally be emphasized that our sources are not always exact in their calculation of time if we do not accept a slight deviation. The number of years for which a particular king reigned or an elder kept the *Vinaya* is given in rounded-off numbers in our records, with months and days being ignored. A deviation of a couple of years one way or another cannot be denied in a calculation involving about 100 years or so. Thus, 397 BCE may only be taken as a rough approximation to the year in which the Buddha expired. Some of the scholars who initially played an important role in popularizing the long chronology have now reverted to the short chronology, thus adding to its growing popularity. For instance, André Bareau, shortly before his death, in his "The Problem Posed by the Date of the Buddha's Parinirvana," revised his position and proposed that "in placing the Parinirvana of the Blessed One around 400, with a margin of twenty years added or deduced from this date, we would probably not be very far from the historical truth, which unfortunately remains inaccessible to us with more precision" (1953).

References and Further Reading

Bareau, A. "La date du nirvana." *Journal Asiatique* 241 (1953): 27–62.

Bareau, A. "The Problem Posed by the Date of the Buddha's Parinirvana." In *When Did the Buddha Live?* (pp. 211–19). Edited by H. Bechert. Delhi: Sri Satguru Publications, 1995.

Beal, S. *Si-yu-ki: Buddhist Records of the Western World*, Vol. II. London: Trubner, 1906.

Bechert, H., ed. *The Dating of the Historical Buddha/Die Datierung des historischen Buddha*, 2 vols. Göttingen, Germany: Abhandlungen der Akademie der Wissenschaften, 1991–92.

Bechert, H., ed. *When Did the Buddha Live?* Delhi: Sri Satguru Publications, 1995.

Bhattasali, N. K. "Mauryan Chronology and Connected Problems." *Journal of the Royal Asiatic Society*, Pt. II (1932): 283.

Bühler, G. "Three New Edicts of Ashoka." *Indian Antiquary* 6 (1877): 149–60.

Eggermont, P. H. L. *The Chronology of the Reign of Asoka Moriya*. Leiden, Netherlands: Brill, 1956.

Eggermont, P. H. L. "New Notes on Ashoka and His Successors II." *Persica* 4 (1969): 97.

Frauwallner, E. *The Earliest Vinaya and the Beginnings of Buddhist Literature*. Rome: Is. M.E.O., 1956.

Geiger, W., ed. *The Mahavamsa or the Great Chronicle of Ceylon*. Trans. W. Geiger and M. H. Bode. London: Pali Text Society, 1912.

Halbfass, W. "Early Indian References to the Greeks and the First Encounters between Buddhism and the West." In *When Did the Buddha Live?* (p. 205). Edited by H. Bechert. Delhi: Sri Satguru Publications, 1995.

Law B. C., ed. and trans. *The Chronicle of the Island of Ceylon or the Dapavamsa*. Colombo: Ceylon Historical Journal, 1958.

Lienhard, S. "A Brief Note on the Date of the Historical Buddha and Classical Poetry." In *The Dating of the Historical Buddha/Die Datierung des historischen Buddha* (Vol. 1, p. 196). Edited by H. Bechert. Göttingen, Germany: Abhandlungen der Akademie der Wissenschaften, 1991.

Max Müller, F. "The True Date of Buddha's Death." *The Academy* (March 1884): 153.

Oldenberg, H. "Die Datierung der neuen angeblichen Ashoka-Inschriften." *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 35 (1881): 472–76.

Pachow, W. "A Study of the Dotted Record." *Journal of the American Oriental Society* 83, 3 (1965): 342–45.

Rhys Davids, T. W. "The New Ashoka Inscriptions." *The Academy* (July 1877): 37.

Rhys Davids, T. W., and J. E. Carpenter, eds. *The Dagha Nikaya* (The Dialogues of the Buddha), 3 vols. Translated by T. W. and C. A. F. Rhys Davids. London: Pali Text Society, 1899, 1910, and 1957.

Sarao, K. T. S. "Urban Centres and Urbanization as Reflected in the Pali Vinaya and Sutta Pitakas." Ph.D. thesis submitted to the University of Cambridge, 1989.

Schmithausen, L. "An Attempt to Estimate the Distance in Time between Ashoka and the Buddha in Terms of Doctrinal History." In *The Dating of the Historical Buddha/Die Datierung des historischen Buddha* (Vol. 2, p. 143). Edited by H. Bechert. Göttingen, Germany: Abhandlungen der Akademie der Wissenschaften, 1992.

Stein, O. "The Coronation of Candragupta Maurya." *Archiv Orientalni* 1 (1932): 368.

Takakusu, J. "Pali Elements in Chinese Buddhism: A Translation of Buddhaghosa's *Samantapasadika*, a Commentary on the *Vinaya*, Found in the Chinese *Tripitaka*." *Journal of the Royal Asiatic Society* (1896): 436.

Thomas, E. J. "Theravādin and Sarvāstivādin Dates of the Nirvāna." In *B. C. Law Volume* (Vol. 2, pp. 18–22). Delhi: Poona, 1946.

Winternitz, M. *A History of Indian Literature*, Vol. II. Calcutta: University of Calcutta, 1933.

Index

POPULAR CONTROVERSIES

in

WORLD HISTORY

STEVEN L. DANVER, EDITOR

Popular Controversies in World History

INVESTIGATING HISTORY'S INTRIGUING QUESTIONS

Volume Three
The High Middle Ages to the Modern World

Steven L. Danver, Editor

ABC-CLIO

I

North American rune stones point to extensive exploration by the Norse of North America.

PRO   Justin Corfield
CON   Harald Haarmann

PRO

There have long been sagas that recorded the Vikings—or Norsemen, as they were known at the time—sailing across the Atlantic Ocean. They certainly established settlements in Iceland and Greenland, including the west coast of Greenland. There are also two Icelandic sagas, the Greenlander's saga and Erik's saga, recording the locating of what became known as Vinland. Both of these record how Erik the Red who, outlawed from Iceland in the late 10th century, sailed for Greenland and built a colony there. In ca. 985–986, a Norse leader named Bjarni, whose father had sailed with Erik the Red, while at sea, was blown off course by a violent storm and ended up on the coast of a place that had neither snow nor mountains. The Norsemen decided not to land because Bjarni wanted to reach Greenland, and they managed to return there. However, 15 years later, the son of Erik the Red, Leif Eriksson "the Lucky," sailed to a place identified by many as the coast of America. The first place where he landed was called the Helluland, or "Land of Flat Stone," and then he and his men headed south to what he named Markland, or "Land of Woods." They later landed at a spot that had beaches of "silver" sand and decided to stay ashore for a few days. It was during this time that they found grapes that one Norseman, who had served in Turkey (probably in the Varangian Guard), identified correctly, leading the Norsemen to name the place Vinland.

By studying the various sea currents, and given that the Norse settlements on the west coast of Greenland were located at Godthab and Sandnes, it has been suggested that Helluland could be the southern part of Baffin Island, with Markland being Labrador and Vinland being the region around the island of Newfoundland. This theory meant that the Norsemen probably did not reach any part of the modern United States. Brief mention should also be made of the "Vinland map," which was discovered in 1965 and marked identifying Helluland, Markland, and Vinland. The map became the subject of an international conference at the Smithsonian Institute in Washington, D.C., in 1966, and it is now believed to have been forged, perhaps as recently as the 1920s.

I
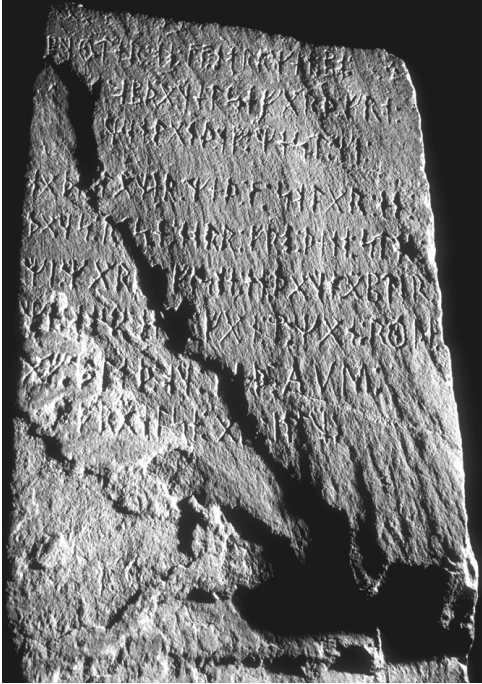
The Kensington Runestone, found in 1898 near Kensington, Minnesota by a farmer named Olof Ohman, has been a source of continuing controversy over its origins. (Richard Hamilton Smith/Corbis)

Although the visits of Bjarni and Leif Eriksson have been well known to historians for many years, there have always been queries about whether any subsequent Viking expeditions reached North America. With mounting evidence from the sagas that some Vikings in the 10th century had given the name Vinland to part of North America, there was no physical evidence found for the Vikings being in the Americas until 1898, with the discovery of what became known as the Kensington Runestone.

In 1893 a Danish archaeologist had shown that it was possible for a Viking ship to reach America, and there was then renewed interest in the subject. Only five years later, a rune stone was found in 1898 on a farm some three miles from the village of Kensington, in the southwest corner of Douglas County, in west-central Minnesota. As a result of where it was found, it became known as the Kensington Runestone. The man who found it was Olof Ohman, a Swedish-born farmer who had migrated to the United States in 1879 and was living on the farmstead with his wife Karin and their children.

The place where the stone was found was near the crest of a small rising piece of land above the wetlands. According to the version by Ohman, it was tangled in the root system of a poplar tree as though it had been placed near the site of the tree, and the roots had wrapped themselves around it. Olof's eldest son, also called Olof, aged 10 at the time, uncovered the stone and spotted some lettering on it and, apparently, believed that it was some kind of "Indian almanac." The stone was then pulled out of the ground. It was 30 inches tall, 16 inches wide, and 6 inches thick (76 by 41 by 15 cm), and weighed about 200 pounds (90 kg). It was not long after its discovery that Olof Ohman, the boy's father, began to realize what the stone might represent. From the 1870s, schoolchildren in Norway and Sweden had been taught about runes as a part of the syllabus covering Scandinavian heritage.

The Kensington Runestone was then taken to a local bank where it was put on display. There is certainly no evidence that the Ohmans tried to make any

money from the discovery, and this is one of the key points to be raised in favor of the discovery being genuine. A transcription of the stone was quickly made, which was later found to be inaccurate, and it was sent to Olaus J. Breda, who was a professor of Scandinavian languages and literature at the University of Minnesota. Although there are some queries about his linguistic knowledge, he had been born in Norway 45 years earlier and was clearly the main expert in Minnesota at the time, having previously taught in Iowa.

Olaus Breda translated the text and immediately declared that the stone was a clear forgery, and many other linguists concurred. It was subsequently translated by Hjalmar Holand from the lettering, or runes, as: "8 Geats (South Swedes) and 22 Norwegians on acquisition venture from Vinland far to the west We had traps by 2 shelters one day's travel to the north from this stone We were fishing one day. After we came home found 10 men red with blood and dead AVM (Ave Maria) Deliver from evils!" On the other side there was an inscription: "I have ten men at the inland sea/lake to look after our ship 14 days travel from this wealth/property. Year of Our Lord 1362" (Holand 1909: 15).

It was not long after its discovery that the stone started to gain some political notoriety. By 1898 there were considerable political problems between Norway and Sweden, with Norway finally achieving its independence in 1905. As the stone described a joint Norwegian-Swedish expedition, some Norwegians started to claim that the stone was actually a Swedish hoax, with Swedes keen on proving a long historical link between the two countries. Being found by a Swedish farmer further emphasized their point.

The Kensington Runestone was then sent to Northwestern University in Chicago, and later returned to the Ohmans, with most scholars beginning to believe it was a hoax. However, in 1907, Hjalmar Rued Holand bought the stone from the Ohmans for a reported $10. Holand believed the stone was genuine and spent most of his life researching it.

Hjalmar Holand had been born near Oslo, Norway, on October 20, 1872, and when he was 12 he migrated to the United States with his parents Johan Olson Holand and Maren (née Rued). In 1898 he graduated from the University of Wisconsin-Madison and gained his master's degree the following year. He then started farming near Ephraim, Wisconsin, before becoming very interested in the Kensington Runestone. Although some books describe him as a furniture salesman, Holand is described in the 1910, 1920, and 1930 censuses as a farmer or a fruit farmer.

To try to ascertain whether or not the stone was genuine, Holand and others made a number of investigations. The Ohmans certainly did not profit from their find, although they did enjoy the early local publicity. When they first found it, they felt it was a curiosity. Their exhibition of it in a local bank certainly indicates this, as well as their treatment of it when it returned from Northwestern University—they kept it in a shed until Holand bought it. Although there is a

popular story that the family used it for straightening out nails or as a stepping-stone into the granary, neither of these stories has been confirmed, and even if true, they only further emphasize the family attaching no huge value to it.

One of the initial areas for investigation was confirming the exact location where the stone was found and why it was found in Kensington, Minnesota. The land was not settled in modern times until 1858, and even then with the Dakota War of 1862, there were relatively few initial settlers, although the land had been taken from the Native Americans by the late 1860s. The tree under which the stone was found was cut down before it could be studied and its age ascertained, but other nearby poplars were dated as being about 40 years old in the 1890s, although one person who had seen the original poplar did feel that it might have been only 12 years old.

But the question that remained at that point unanswered was why the stone was found in Minnesota in the first place. By the 1900s, many people were beginning to believe that the Vikings had arrived in North America in medieval times, but the concept of them sailing up the Saint Laurence River and through the Great Lakes to reach Minnesota, had not been considered. To reach the spot, the Vikings would then have had to have gone south along the Minnesota River before going north along Chippewa River. It was an extremely long journey from the Atlantic Ocean. It is not altogether impossible, but this led to many historians querying the validity of the stone solely on account of its location.

Holand started investigating the nearby waterways and at Cormant Lake, Becker County, also in Minnesota, where he found boulders that had triangular holes in them. He compared them to similar ones found off the coast of Norway, which were used to anchor boats in the 14th century. It appeared that the stones in Minnesota had been fashioned with chisels rather than drills. Also, in 1871 at Climax, Minnesota, a firesteel had been found that matched a similar specimen held at the museum at Oslo University, Norway. There also remain two other scenarios. Because many Native American tribes had inhabited Minnesota, there has to remain the possibility that the stone might have been moved by them between its original erection on the coastline and its discovery by Ohman, or it might have been taken there by an early European explorer. Mention should also be made of the Viking Altar Rock at Sauk Centre, Minnesota: a rock that is 27 feet (8.2 meters) long and 17 feet (5.2 meters) wide, it has two sets of holes about three feet from the base, which may have held an altar shelf, with holes at each end to support halberds from which a canopy could be hung. Although this altar was rededicated in August 1975 and is now a part of the Trail of the Vikings in Minnesota, even if it were not genuine, it would have no real bearing on the Kensington Runestone.

Newton Horace Winchell, a geologist who was also a member of the Minnesota Historical Society, and George Tobias Flom, a linguist from the Philological Society of the University of Illinois, both made an extensive study of

the rune stone in 1910. They were the last ones to make a detailed study based on a physical analysis for the next 88 years. N. H. Winchell (1839–1915) was the state geologist of Minnesota from 1872 until 1900 and professor of geology and mineralogy at the University of Minnesota from 1873 until 1900. He was also the founder and editor of the journal *American Geologist* from 1888 until 1905. His son was also a prominent geologist. G. T. Flom (1871–1960) was born in Wisconsin, the son of Norwegian parents, and became a professor of Scandinavian languages and acting professor in English philology at Iowa State University from 1900 until 1909, then professor of Scandinavian languages at the University of Illinois from 1909 until 1927, and professor of Scandinavian languages and English philology, writing many learned papers on paleography and Norwegian medieval texts. Winchell sought out the evidence as to where the stone was found and largely drew a blank, with Flom querying whether the inscription used more modern language than he felt might have been the case in the 14th century.

To try to verify the authenticity of the Kensington Runestone, Hjalmar Holand took the stone with him to Europe in 1911 and showed it to scholars at the Norse Millenary Congress held at Rouen, France, in that year. He did not have much success there, with scholars such as Charles H. Haskins claiming that a cursory examination of the stone showed that the rune had been cut recently and other scholars claiming that there were linguistic mistakes.

The linguistic analysis of the Kensington Runestone, started by Breda, has been intense, and much of it rests on the use of the letter "*j*." Because this letter was not used at the time of the runestone's supposed origin, most linguists were quick to denounce the stone a as forgery. However, Robert A. Hall Jr., emeritus professor of Italian language and literature at Cornell University, claimed there might not be as many linguistic problems as others had thought. His account, *The Kensington Rune-Stone is Genuine: Linguistic, Practical, Methodological Considerations*, was published in 1982. Hall studied the normal variances in dialects in Old Swedish and suggested that the linguistic evidence was inconclusive, which made other aspects of the stone crucial in deciding its authenticity. This led to an engineer, Richard Nielsen, from Houston, Texas—his work being heavily inspired by that of Hall—studying the stone and in 1983 he argued that what had been interpreted as the letter "*j*" was, in fact, the letter "*l*," which was used in the 14th century, and although rare, it was quite in keeping with its use on the Kensington Runestone. These linguistic analyses, however, were merely opinions. The people who carved the runes on stone could easily have made a mistake or misspelling—it happens to this day even on official plaques and gravestones—and there are manuscripts found regularly around the world that contain new words or ideas previously not thought to have been used at the time.

As to the finding of the stone, some historians such as Theodore C. Blegen had long believed that Ohman or others had forged the stone—there were many

Scandinavian settlers there. This theory gained support when an audiotape was transcribed in 1976, on which another farmer, Walter Gran, claimed that his father, John Gran, had confessed, in 1927, to having made the stone with Ohman. However, although this new claim received some publicity, Walter Gran never seems to have said anything more about it before his death some years later, and some locals claimed that there was jealousy between the Grans and the Ohmans, which might have been the real motivation for the release of the tape.

However, Blegen, in his *The Kensington Rune-Stone: New Light on an Old Riddle*, published by the Minnesota Historical Society in 1968, offers a different suggestion. He claimed that a Swedish schoolteacher named Sven Fogelblad (1812–1897), who grew up in Sweden and immigrated to Minnesota and was a friend of the Ohmans, may have been the source of the carving of the rune stone. Furthermore, he noted that Fogelblad had graduated from Upsala University and had been a cominister in Sweden with a noted rune expert Claes J. Ljungström, who had written on the subject. It appears that Fogelblad had left the church over dogma differences, but little more is known about his life. He does not appear on any U.S. Census, at least under the surname Fogelblad. And if Fogelblad had been the source of the stone, there remain some queries, the first one being why the stone was only found after Fogelblad died. Also, if, as some historians claim, Fogelblad was the forger, based on being an expert on runes, there has to be some question about why he would make the elementary mistakes these same authors claim are present on the stone.

The issue of the authenticity of the Kensington Runestone continued for many years. By this time Hjalmar Holand had written extensively on the subject. His *History of the Norwegian Settlements* had been published in 1908, soon after he had bought the Kensington Runestone, and his first major work on the subject, *The Kensington Stone*, was published in 1932, followed eight years later by *Westward from Vinland*. His book *America 1355–1364* was published in 1946, and *Explorations in America* was published exactly 10 years later. Indeed, in 1948–1949 the stone was displayed at the Smithsonian Institution, with a photograph of it appearing in *National Geographic* in September 1948. The photograph showed Neil Morton Judd (1887–1976), curator of American archaeology at the Smithsonian from 1919 until 1929, and curator of archaeology there from 1930 until 1949, examining the stone that the caption claims has been declared genuine by recent studies of it.

By this time some written confirmation of the possibility of the Norsemen in North America in the 14th century had emerged. Carl Christian Rafn had written about some of this in the early 19th century, after uncovering a reference from the cartographer Gerardus Mercator (who developed the Mercator projection) in a 1577 letter in which he refers to a geographic text about the Arctic region of the Atlantic, written over 200 years earlier by a man named Jacob Cnoyen. This letter mentions that eight men had returned to Norway from

the arctic in 1364. Accounts of Greenland by Ivar Bardarson, a cleric, also exist from the same year. There were also references from Gisli Oddsson, the bishop of Iceland, who started writing a history of Iceland and other Norse colonies in 1637 and mentioned that some of the colonists in Greenland had rejected Christianity in 1342 and went to North America. There was also a letter by King Magnus II "Eriksson" of Sweden and Norway ordering Paul Knutson to lead an expedition to sail for Greenland, although there is no evidence that this expedition ever set out—and, indeed no evidence that it did not.

These problems over the authenticity of the Kensington Runestone led to academic debate throughout the 1950s and 1960s. S. N. Hagen and William Thalbitzer acknowledged its authenticity in a series of articles. Hagen's account, "The Kensington Runic Inscription," was published by the journal *Speculum* in 1950; and Thalbitzer wrote his account, *Two Runic Stones, from Greenland and Minnesota*, which was published by the Smithsonian Institute in 1951. Holand continued writing about the rune stone with *Explorations in America*, published in 1956, and *A Pre-Columbian Crusade to America*, published in 1962. This last work sadly links the stone with the "Newport Tower" and several other supposed Norse sites that have been totally disproved. Critics have also used these to attack Holand's thesis on the Kensington Runestone. Holand died on August 8, 1963.

Two years before Holand died, Ole G. Landsverk started publishing his thoughts on the Kensington Runestone with his *The Discovery of the Kensington Runestone: A Reappraisal*. In 1974 his *Runic Records of the Norsemen in America* was published in New York, and his "The Kensington Inscription Is Authentic" was published in the *Journal of the New England Antiquities Research Association* in 1981. It was not long before scientific analysis of the Kensington Runestone came up with important new discoveries.

Advances in science and technology led to a much more detailed study of the stone itself from December 1998 onward. Curiously, exactly 100 years after its discovery by Ohman, the 1998 examination was the first detailed physical analysis of the stone since that by N. H. Winchell in 1910. This new study involved sampling of the core of the stone and examination of the stone with a scanning electron microscope. A geologist, Scott F. Wolter, released the preliminary findings at the Midwest Archaeological Conference, held in St. Paul in November 2000.

Wolter said that the evidence of the stone itself suggested that it had been "in the ground" for between 50 and 200 years. He compared the stone with slate gravestones from Maine dating back 200 years—gravestones were used because an exact date of when they were set could be ascertained—and these gravestones showed considerable degradation of the pyrites, whereas the Kensington Runestone showed a total lack of mica on the inscribed surface of the stone. Even though the effects of weathering between Maine and Minnesota might not be exactly the same, Wolter was able to show scientifically that the stone had been buried in the ground long before the first permanent European settlement of that

part of Minnesota in 1858. This could also explain the entangling in the roots of the poplar tree. It certainly placed the stone well before either Sven Fogelblad or Olof Ohman arrived in the United States—indeed before Ohman had been born.

However, the more important information came forward when Wolter examined all the runes on the Kensington Runestone and found that there were a series of dots engraved inside the four runes forming an "*R*" shape. This was similar to dotted runes located in 14th-century gravestones on the island of Gotland, off the east coast of Sweden near Stockholm. This again tended to give proof that the Kensington Runestone was genuine.

In 2001 *Scandinavian Studies* published an article by Richard Nielson that argued that although a number of the runes on the Kensington Runestone were similar to the Dalecarlian Runes (from Dalarna, Sweden), others had no connection to them and could easily be explained by 14th-century dialectic differences. Three years later, Keith Massey and his brother Kevin Massey claimed that the use of "AVM" for Ave Maria was a medieval tradition that would not have been known to the Ohmans or any other forger in the 1890s. Although the Ohmans and Fogelblad might have had access to other information, the religious abbreviation would not have been known to them.

In 2004 a new book added to the body of information suggesting that the Kensington Runestone was genuine. Written by Alice Beck Kehoe, an archaeologist, her book *The Kensington Runestone: Approaching a Research Question Holistically* set out to investigate one area that had hitherto not been the subject of research. She unearthed some Native American traditions that included references to an ancestral hero known as Red Horn who managed to encounter "red-haired-giants." She was also able to trace signs of a tuberculosis epidemic among some Native American tribes in medieval times and 19th-century accounts of Native Americans with "blond hair." Although these oral traditions are not conclusive, more weight is placed on them now than was the case a century earlier. The Kensington Runestone is now in the Runestone Museum in Alexandria, Minnesota.

Although the controversy over the Kensington Runestone continues, it is now accepted by historians that the Norsemen did arrive off the coast of North America. In 1960–1961, Norwegian archaeologists Helge and Anne Ingstad, working in Newfoundland, Canada, uncovered what they believed was a Norse settlement located at L'Anse aux Meadows. Since then they and others have found many more traces of Norse civilization in remote parts of the west coast of North America.

L'Anse aux Meadows is on the northern tip of the island of Newfoundland, and the Ingstads, working there, founded a settlement that was proven to consist of at least eight buildings. These included a forge, a smelter, and a lumberyard, which could have provided enough wood for building a ship. The largest house there was 28.8 meters by 15.6 meters, and it had several rooms. A detailed archaeological dig on the site uncovered sewing and knitting tools that indicated

---

The Runestone Museum

Located in Alexandria, Minnesota, near where Olaf Ohman found the Kensington Runestone, the Runestone Museum not only celebrates the evidence of Nordic exploration carried by the Runestone, but also pays homage to Native American cultures, the Scandinavian population of the region, and historic Fort Alexandria. Just in case visitors miss the town and its significance, a 28-foot statue of a Viking, affectionately called "Big Ole," sits in downtown Alexandria, proudly proclaiming the town as the "birthplace of America." A 40-foot Viking ship named the "Snorri" educates visitors on how the Vikings made their way to North America, while a wildlife exhibit and a replica of the 1862 fort teach adults and children alike about the region's long history.

---

that women might have been present at the site. The finds quickly attracted much attention throughout the world, with a Norwegian adventurer-turned-archaeologist Helge Ingstad and his wife working on the site. Ingstad wrote about them in the *National Geographic* in November 1964, and then in his book *Westward to Vinland* (1969). The site at L'Anse aux Meadows is now accepted by most as being a major Norse archaeological discovery, although Erik Wahlgren (2000), who has long denounced the Kensington Runestone as a hoax, has argued that the main Norse settlement in North America was farther south than L'Anse aux Meadows. It should be mentioned that Helge Ingstad also felt that the Kensington Runestone was not genuine, but this, and Wahlgren's work, was before the recent scientific tests—Wahlgren died in 1990, and Helge Instad died in 2001, aged 101.

Although the Kensington Runestone is not the only Norse "stone" in North America, it is certainly far more important than the others. Some writers have associated the Heavener Runestone with the Kensington Runestone. Found at Heavener, Oklahoma, during the 1890s, if Norsemen placed it in Oklahoma, it shows that they managed to navigate the Mississippi River, the Arkansas River, and the Poteau River to the site where it was found. Although possible, it seems an unlikely location, but this has created much interest in Oklahoma, yet only became famous after Frederick Julius Pohl wrote about it in his book *Atlantic Crossings before Columbus*, published in 1961.

There are a number of theories about the inscription on it "GNOMEDAL," suggesting it might mean the Valley of Glome ("Glome Dal"), or even refer to the date "November 11, 1012," as suggested by Alf Monge in 1967. The Heavener Runestone was still officially recognized as genuine by the Oklahoma Historical Society in 1959, although other historians feel that its location makes it unlikely to have been Norse in origin. It became famous internationally with the novel *Runestone*, by writer Don Coldsmith, published in 1995. Once again,

as with the Kensington Runestone, it is possible that Native Americans or later settlers might have moved the Heavener Runestone.

Another rune stone, also found in Oklahoma, the Shawnee Runestone, spells out "Medok," which has been taken by some to refer to the Welsh Prince Madoc, who was said to have sailed for North America in 1170, or, according to Alf Monge, is a cryptopuzzle giving the date "November 24, 1024." The Oklahoma filmmaker Jackson Burns, in a film on the Heavener Runestone, portrayed the Shawnee Runestone as a hoax made up about 50–60 years ago, and the two other rune stones found near that of the Heavener Runestone as also being hoaxes, but that the Heavener Runestone and another known as the Poteau Runestone might be genuine. However, they have all been subjected to much less interest than the Kensington Runestone, which, after recent scientific tests, is now believed by many to be genuine.

References and Further Reading

Blegen, Theodore C. *The Kensington Rune Stone, New Light on an Old Riddle*. St. Paul: Minnesota Historical Society, 1968.

Hagen, S. N. "The Kensington Runic Inscription." *Speculum* 25 (1950): 321–56.

Hall, Robert A. Jr. *The Kensington Rune-Stone Is Genuine: Linguistic, Practical, Methodological Considerations*. Columbia, SC: Hornbeam Press, 1982.

Holand, Hjalmar R. *America 1355–1364: A New Chapter in Pre-Columbian History*. New York: Duell, Sloan and Pearce, 1946.

Holand, Hjalmar R. *Explorations in America before Columbus*. New York: Twayne, 1956.

Holand, Hjalmar R. "An Explorer's Stone Record which Antedates Columbus." *Harper's Weekly* 53 (October 9, 1909): 15.

Holand, Hjalmar R. *The Kensington Stone: A Study in Pre-Columbian American History*. Ephraim, WI: Private printing, 1932.

Holand, Hjalmar R. *A Pre-Columbian Crusade to America*. New York: Twayne, 1962.

Holand, Hjalmar R. *Norse Discoveries and Explorations in North America 982–1362: Leif Eriksson to the Kensington Stone*. New York: Dover, 1969.

Holand, Hjalmar R. *Westward from Vinland: An Account of Norse Discoveries and Explorations in America, 982–1362*. New York: Duell, Sloan and Pearce, 1940.

Hughey, Michael W., and Michael G. Michlovic. "Making History: The Vikings in the American Heartland." *International Journal of Politics, Culture and Society* 2, 3 (1989): 338–60.

Ingstad, Helge. *Westward to Vinland: the Discovery of Pre-Columbian Norse House-Sites in North America*. New York: St. Martin's Press, 1969.

Kehoe, Alice Beck. *The Kensington Runestone: Approaching a Research Question Holistically*. Long Grove, IL: Waveland Press, 2004.

Landsverk, Ole G. *The Discovery of the Kensington Runestone: A Reappraisal.* Glendale, CA: Church Press, 1961.

Landsverk, Ole G. "Does Crytogrammic Analysis Reveal Pre-Columbian Voyages to America?" *Biblical Archaeology Review* 6: 2 (1980): 54–55.

Landsverk, Ole G. "The Kensington Inscription is Authentic." *Journal of the New England Antiquities Research Association* 15: 3 (1981): 58–62.

Landsverk, Ole G. *Runic Records of the Norsemen in America*. New York: E. J. Friis for Twayne, 1974.

Morison, Samuel Eliot. *The European Discovery of America: The Northern Voyages*. New York: Oxford University Press, 1971.

Nielsen, Richard. "Response to Dr. James Knirk's Essay on the Kensington Runestone," *Scandinavian Studies*, 72:1 (Spring 2001), 2-75.

Pohl, Frederick J. *Atlantic Crossings before Columbus*. New York: Norton, 1961.

Quaife, Milo M. "The Kensington Myth Once More." *Michigan History* 31 (June 1947): 129–61.

Rafn, Carl Christian. *America Discovered in the Tenth Century*. New York: W. Jackson, 1838.

Reman, Edward. *The Norse Discoveries and Explorations of America*. Berkeley: University of California Press, 1949.

Thalbitzer, William. *Two Runic Stones, from Greenland and Minnesota*. Washington, DC: Smithsonian Institution, 1951.

Wahlgren, Erik. *The Vikings and America*. London: Thames and Hudson, 2000.

CON

The thrust of this chapter, that the so-called Kensington Runestone is a forgery, has been argued against by the stone's apologists for a century, but such arguments have demonstrated inherent flaws. For any assessment of the role of runic inscriptions as evidence of Norse explorations in North America to be comprehensive, it does not suffice to analyze runic material in isolation. It is the socioeconomic and cultural contexts of living conditions during the Viking age that provide clues to the motivation for using the runic script by the Norse communities in the North Atlantic region. An analysis of these conditions is crucial for understanding why rune stones were not erected in Greenland or on the American mainland, and why runic material remained scarce in that part of the world.

In the discourse about the significance and extension of Norse explorations of North America, manifold facets of a mixture of facts and fiction are revealed. We can learn valuable lessons about the reactions of the human mind to the soberness of scientific findings and to the temptation of wishful thinking. Arguably, the personal history of anyone who participates in the debate has a bearing on his or her attitude toward the Vikings, their quests, and their culture. In the minds of central and southern Europeans, Viking seafaring and colonization may have the value as an academic subject, and Viking mythology and narratives may be themes for entertainment.

For someone from Scandinavia, though, this takes on a different meaning. Norwegians and Icelanders are very interested in their family histories, and they celebrate ancestors who once set out to sea, as adventurers, to colonize Greenland. Having a Greenlander among one's kinfolks is a matter of great pride. The historical retrospective is also sentimental for many Scandinavian Americans. The world of archaeologists who deal with finds from the Viking age and the views of romanticizers of the Norse explorations are colliding to the effect that "a gulf has emerged between scientific and popular opinion of what the Vikings did and did not do in North America. The contested ground between scholarship and popular or folk culture has been a constant and problematic feature of the social side of American archaeology" (Wallace and Fitzhugh 2000: 374).

History is never fixed because the time distance that separates us from past events is continuously floating, and every epoch has its own zeitgeist that sets standards and conventions of historical retrospective. In other words, every epoch and its zeitgeist forge the conceptual molds of our cultural memory. With the rise of nationalism in the 19th century, the pride in a nation's cultural heritage grew to unprecedented scale. Those people from Scandinavia who, some 100 or 150 years ago, immigrated to North America, brought with them the national pride in their cultural heritage, above all the glorified Viking past, which they continued to cultivate in their new homeland.

Between 1850 and 1875, about 370,000 people from Scandinavian countries immigrated to the United States. In the 1880s, immigration from Scandinavia increased to almost half a million people. The highest rates for immigration are recorded for the decade between 1901 and 1910 (Gjerde 1995: 87). This considerable portion of the population in the United States made a difference when it came to the shaping popular opinions about the pre-Columbian history of North America. The offspring of immigrants of Scandinavian descent were educated in the spirit that had shaped the cultural memory of their parents and grandparents. From one generation to the next, Viking history became increasingly sentimentally mystified, as an ingredient of people's self-identification.

The Scandinavian spirit manifested itself in the popular discussion about the Norse quests in North America at an early time, in the 1850s, to the amazement of contemporaries. Since American history is much younger than that of other continents, there seems to have been a particular interest in pre-Columbian

events: "[I]t appears refreshing to light on any class of facts which promises to lend a ray of antiquity to our history" (Schoolcraft 1853–57, I: 109). The arrival of the Vikings on the east coast of North America predates the "discovery" of the continent by Columbus by almost five centuries. The perception of time may vary considerably in a cross-cultural comparison. The date of 1000 CE as the beginning of the Norse presence in North America corresponds to the period of the early Middle Ages in Europe. This horizon of time would not seem long to many Europeans, but, for Americans, it signals a remarkably "remote" past.

The memory of the pre-Columbian past was constantly nourished by finds, both historical and fanciful, that seemed to be somehow related to the Viking era, especially objects that were allegedly inscribed, or in the words of a contemporary, "hardly a year passes unsignalized by the announcement of the discovery of tablets of stone or metal, bearing strange and mystical inscriptions" (Squier and Davis 1848: 274). The speculative mind of American historians readily clung to these disparate finds that were woven into a popular narrative about the heroic explorations of North America by the Vikings. In 1874, Rasmus B. Anderson, professor of Scandinavian languages at the University of Wisconsin, published his *America Not Discovered by Columbus*, which not only appealed to readers with an academic background but also became popular. An event that caused a sensation was the arrival of the "Viking," a replica of the historical Gokstad vessel that had crossed the Atlantic Ocean, in Chicago in 1893.

In the 19th century, speculations of research added a facet of global prestige to the theme of Viking explorations and cultural achievements, highlighting the high age of the runic script. The 19th century was still a time when students of the history of writing had freedom to speculate about the historical relationships among ancient writing systems because reliable dating methods of artifacts and cultural strata did not exist at that time. Karl Faulmann's 1880 work, *Illustrirte Geschichte der Schrift*, the first universal history of writing ever compiled, engages in chronological speculation by assuming that the system of Germanic runes was the oldest script of humankind and that the Mesopotamian and ancient Egyptian traditions were inspired by the runes in both the outer form of their signs and their manner of stylization.

Faulmann's assumption about the runic system being the mother of all other writing systems was certainly inspired by a Euro-centric bias, a common attitude of the time (Lambropoulos 1993). His ideas spread widely and became popular, spurring numerous attempts to renew Faulmann's claim up to the present. It took decades before the relative age of the runes could reliably establish the origins, which are dated to the first century CE. As for the time frame of the emergence of runic writing, the flair of high age dissolves in a comparative view:

Compared to ancient classical scripts, the old Germanic runes are a young and unsophisticated relative from a remote corner of the world. Runes may

---

The Kensington Runestone and the Holy Grail

At least one modern geologist, Scott Wolter (2005), believes and has written about the authenticity of the Kensington Runestone. Not only that, but Wolter also believes that the rune stone was carved by members of the medieval military order, the Knights Templar, after their voyage to North America over 100 years before Columbus sailed. The Knights Templar fled Europe after the pope disbanded their order in the 1300s, and some have argued that they took all they could carry on their ship, including important relics such as the Holy Grail, and sailed to America. Though Wolter believes the rune stone is a document laying claim to the Mississippi River Valley, and has no direct ramifications for grail lore, if his theory is true, then the grail might just be buried somewhere in the region as well.

---

be called young, since, at the time when they first appear, almost all major writing systems in the world were already invented. (Williams 2004: 262)

This means that the runic script was a latecomer to the world of literacy, as was the Celtic Ogham.

Seeing Norse Explorations of North America in the Archaeological Record

In the consciousness of 19th-century Americans, their country was strewn with remainders of the Norse presence on the American continent. The list of the artifacts is long, which were thought of as proving the existence of Viking settlements, of Norse explorations of the coastal areas, and as far as the Midwest. But most items on that list do not stand the scrutiny of scientific verification. Many artifacts have revealed their true nature as fakes, and the ruins of buildings have proven to be of an age unrelated to the Viking era. This is true for the Kensington stone from Minnesota, the Newport Tower, the Heavener rock carvings in Oklahoma, the Beardmore relics from Ontario, the Spirit Pound Runestones from Maine, and others.

In the 1960s, the ruins of an undisputedly Norse settlement could be identified. The site is L'Anse aux Meadows in Newfoundland, which became the first United Nations Educational, Scientific, and Cultural Organization World Heritage site. No other settlements dating to the Viking era have been found in North America, which means that other campsites mentioned in the sagas have so far not been identified on the American mainland.

The interconnection of the Norse settlement at L'Anse aux Meadows with the Viking world can be demonstrated for a diagnostic item of Norse culture: tools made of jasper. The material of the tools found in Newfoundland matches jasper sources from Iceland and Greenland (Smith 2000). Another item of

unequivocal Norse fabric is a coin, a Norse penny, which was unearthed in Maine (the Goddard site on the northeast coast of Penobscot Bay). This piece, minted in Norway between 1065 and 1080, is not evidence for Norse presence in the area but rather of trade relations between the North (i.e., Newfoundland and Labrador) and the South (i.e., areas of Massachusetts and Connecticut) (Cox 2000). The archaeological record of the Goddard site includes Dorset tools from Newfoundland and Rhode Island jasper.

A find that may also be related to the pre-Columbian trade network on the east coast are three bronze pieces (so-called Thor's hammers) from Niantic in Connecticut. The authenticity of these metal objects, however, is disputed because of deviant nickel content, an atypical suspension ring, and surprisingly little corrosion for such old artifacts (Wallace and Fitzhugh 2000: 378).

The Vinland Voyages According to the Icelandic Sagas

The sagas that were composed in Iceland in the Middle Ages, recording the quests of the Vikings in Greenland and North America, are the earliest accounts of the explorations of the North Atlantic by Europeans. The "Greenlander saga" and the "Erik the Red saga" both contain some geographic and ethnographic information relating to the areas that were visited or settled by Vikings. These two sagas form a group that is known as the "Vinland sagas." These sagas, recorded in the Old Icelandic language and written in Latin alphabet, are transcriptions of stories and folktales that were handed down from one generation to the next. In the process, events and narrative themes were subject to various changes. For example, in the "Greenlander's saga," Leif Eriksson is credited as the discoverer of Vinland, and there are accounts of four expeditions. In the "Erik the Red saga," these four expeditions are integrated into one, Leif's role is marginalized, and Thorfinn Karlsefni is hailed as a great explorer. It is therefore problematic to perceive the contents of the sagas as chronicles, which they are not, and the accounts as reports about individual expeditions.

The general fact remains, though, that the Vikings explored the east coast and some islands of North America to some extent. Given their character as narratives rather than chronicles, the Icelandic sagas contain very few exact descriptions of the places and landmarks where Vikings made their landings, set up camps, or founded settlements. Among the places that are given names in the sagas are Straumfjord (Stream Fjord), Straumsey (Stream Island), Hóp (Tidal Pool), Bjarney (Bear Island), Furdhustrandir (Wonder Strand), Kjalarnes (Keel Point), Krossanes (Cross Point), and some others. Three names of geographic areas are mentioned: Vinland, Helluland, and Markland. Since the first half of the 19th century, scholars, amateur historians, and dilettantes have tried to solve the mysteries surrounding the geographic descriptions in the sagas, the first being the Danish philologist Carl Christian Rafn (1795–1864), who

published his "Antiquitates Americanae" in 1837. However speculative most of the identifications may be, they are all confined to a narrow region in the northeast of the North American continent.

The assumption has been made that the site of L'Anse aux Meadows may be identified as either Leifsbudhir (Leif's Camp) or as Straumfjord. According to the context of the saga account, three crews (with men and women from Iceland and Greenland) founded and inhabited the site. Most probably, in this settlement, the first white American born was Snorri, son of Leif and Gudrid. Most of the inhabitants, though, were men and only a few women. There is no archaeological evidence for family households or communal life. "The settlement was not a normal colonizing venture but a base for exploration and exploitation of Vinland's resources farther south" (Wallace 1993: 378). The commodities and material that the Vikings found in Vinland, lumber, furs, walnuts, and grapes, were much appreciated. There was not enough hardwood lumber in Greenland for shipbuilding. The walnuts (i.e., the local subspecies in North America, the butternuts) grew in the Saint Lawrence River Valley, as well as wild grapes.

The site of L'Anse aux Meadows was a seasonal settlement, which means that those who stayed at the site during the summer season arrived in late June and left by mid-September at the latest. The route that was taken went from Brattahlid, the eastern settlement near the southern tip of Greenland, along the shores of Baffin Island (the Helluland of the sagas) and Labrador (Markland) to Newfoundland. That distance is about 2,000 miles (or some 2,300 km, respectively). The voyage lasted between two to four weeks, depending on the winds. The seasonal occupation of Straumfjord may have been between two to three months, from mid-June to mid-September, max. If there were people who lived at the site all year round, their number must have been small. Probably, those who stayed at L'Anse aux Meadows were carpenters who selected lumber and prepared material for the construction of seafaring vessels. The summer season, then, was the time for teams of skilled craftsmen to build new ships that then sailed to Greenland in fall. Some timber would have also been transported, as construction material, to the Norse colonies in Greenland.

The number of occupants of L'Anse aux Meadows is estimated at 40 to 50. This was about 10 percent of Greenland's total Norse population, which ranged between 400 and 500 (Lynnerup 1998: 115). The Icelanders who participated in the Vinland voyages were always a minority. Whatever interests the Vikings had to explore in Vinland, they neither had the manpower nor the resources to keep up a chain of camps along the coast. What was so typical of the colonization of Iceland in the 9th and 10th centuries, the so-called *landnám* ("occupation of the land for settlement"), never happened in Vinland. When it comes to the question whether the Vikings engaged in large-scale explorations of inland North America, as far as Minnesota or Oklahoma, this was simply not possible, given the limited resources they had, compared to what would be required for such expeditions.

L'Anse aux Meadows, the site of an early 11th-century Viking colony at the tip of the Great Northern Peninsula of the island of Newfoundland. The settlement consisted of a number of sod-roofed buildings and included a lumberyard and shipyard. The colony was inhabited for two or three years before it was abandoned. (Dylan Kereluk)

There was a further important factor that would have dimmed the Norse enthusiasm to make inland explorations, if they had any. The Amerindian populations outnumbered the Norse by far, wherever they went, and the encounters with the *skraeling*—as the Native Americans (including the Inuit) were called by the Vikings—were hostile for the most part. The curiosity of Norse explorers was certainly sobered by the perspective of repeated fighting and the high risk of loss of life.

Straumfjord served as a base for the Norse only for a few years. The reason why the Norse abandoned it and the exact date when this happened are unknown. One may assume that the pressure from hostile *skraeling* who lived in the southern parts of Newfoundland or the small parties of seasonal occupants of the site ultimately made the Norse decide not to return there. The halls that had been built as storage facilities and for housing the seasonal occupants were systematically burnt down, either by the Norse before they left or by Newfoundland *skraeling* afterward.

Besides, there was no incentive for the Vikings to explore arable land in North America because, at the time of the Vinland voyages—that is in the 11th century CE—the climate in Greenland was warmer than today, supporting pastures and favoring plant cultivation and the keeping of livestock. As an addition to the agrarian diet, wild caribou provided meat, and rich fishing grounds offered a variety of seafood to the European colonists of Greenland.

Rune Stones in Mainland North America

Asking the question related to the distribution of rune stones is different from asking a question concerning runic inscriptions. The latter is an overarching term, which makes reference to all objects inscribed with runic texts, stationary (i.e., solid rune stones) and mobiliary (i.e., small portable stones, objects made of wood or metal), while the former refers to bigger stones that bear runic inscriptions (Knirk 1993). The great majority of such stones are rocks or boulders that keep their natural form and whose surface was used as a canvas for carving runic letters. In some cases, stones were worked into a rectangular shape and erected.

The statistical record of rune stones shows a marked disproportion in their distribution. A total of 2,307 stones have been surveyed. Of these, 2,057 (89.2 percent) are found in Sweden, 199 (8.6 percent) in Denmark, and 51 (2.2 percent) in Norway. Evidently, the main gravitation of the rune stone tradition lay in Sweden, in the southern part in particular. In light of the wide area with medieval population in Norway, the scarcity of rune stones in that country strikes the eye. One may conclude that erecting rune stones never was a widespread custom in Norway.

It might be for this reason that we do not find any big rune stones, neither in Iceland nor in Greenland where colonists from Norway settled. Rune stones were a tradition that was practiced in the heartland of the Nordic peoples, in Scandinavia, and, to a minor extent, in the British Isles from where some 30 rune stones are known. By the time when the Norse colonization of Greenland began and the first explorations of Vinland took place, the use of runes had already ceased in Britain. The rune stones that were discovered on the North American mainland, under mysterious circumstances, as well as other claimed runic inscriptions are regarded by most scholars as modern forgeries, although some still argue for their authenticity.

The professional skills of the Norse explorers of Vinland included carpentry and metalworking, but not masonry. Those who stayed in L'Anse aux Meadows used specialized tools for shipbuilding and repair and the utensils needed by a metal smith. And it was such tools that the seafarers brought from Greenland. In order to erect a rune stone, the Vinland explorers would have had to rely on someone who was skilled in working big stones and on someone who had training as a literate artisan to execute the design of a runic text on a hard surface. The material that was used to build houses in Greenland were stones, but of relatively small size. Since the custom of erecting stones with inscriptions was not practiced in Greenland, no Greenlander possessed those skills.

If the Norse had felt the need to erect a rune stone for commemorating the discovery of Vinland by Leif, then they would have done that in L'Anse aux Meadows, the pivotal base for all exploration voyages. However, not a single commemorating runic text was carved onto the surface of any of the rocks of which Newfoundland is rich, nor was any stone erected for that purpose.

Besides, no runic material has ever been found at L'Anse aux Meadows, not even small-inscribed objects.

If the Norse did not leave any commemorating texts in Newfoundland, in their major outpost in Vinland, why would those few people who explored other parts of the area erect rune stones elsewhere? Rune stones in Scandinavia were erected in areas with a sizable sedentary population and functioning communities. The Norse colonies in Greenland were fairly underpopulated, and they would have needed more inhabitants to guarantee a normal transition of intergenerational continuity. On the American mainland, there is no archaeological evidence for any colony with a sizable population of Norse descent in pre-Columbian times. In the absence of community life in Vinland, there was no incentive for the Norse to erect rune stones.

Consequently, the claims made by Scandinavian American enthusiasts about the existence of rune stones without associated proof of Norse community life in the area of discovery have raised suspicions among independent scholars about the authenticity of such artifacts. Among the most popular of these runic stones found as isolated pieces, which is outside any cultural context connecting them with Norse archaeological relics, are the Spirit Pond "rune stones" and the Kensington Runestone.

The three "rune stones" of Spirit Pond were found on a beach in Popham (Maine) in 1971. One of the stones showed a map of the area with several place names in runic script. This find aroused much attention in the press, culminating in a story by Calvin Trillin in the *New Yorker* (February 5, 1972). Subsequent investigations demonstrated the stones to be frauds (Wallace and Fitzhugh 2000: 378).

The Dubious Case of the "Rune Stone" from Kensington, Minnesota

None of the other pre-Columbian documents of North America has prompted so much debate about its authenticity as the "runic" text on the Kensington Runestone, which was found in 1898, near Kensington, Minnesota. According to the inscription, a party of some 30 explorers would have reached the region where the stone was erected. The date is given as the year 1362. At that time the Norse settlements in Greenland were in full decline. A drastic climatic change brought colder weather, which made agriculture exceedingly difficult, and in the end impossible, in Greenland. A worsening of socioeconomic conditions can be observed in many parts of northern Europe. Farmsteads were abandoned in great numbers, many of them never to be reinhabited. On top of this, the Black Death took its toll in 1348. "While northern Europe suffered, Norse Greenland died" (McGovern 2000: 327).

Given such unfavorable conditions for the sustenance of the Norse colonies in Greenland, it is absurd to expect contemporary explorations of the American inland. No Greenlander left for America during those times of crisis. The impression of absurdity even increases when taking into consideration that the tradition to

erect rune stones had already been defunct in the heartland of Norse civilization, in Norway, for almost 200 years when the alleged Kensington exploration took place.

We are led to believe that, during a time of irreversible decline of Norse colonies in Greenland and a general worsening of economic conditions in northern Europe, including Scandinavia, a party of some brave explorers—apparently blessed with resources and logistics in abundance—made their way from Newfoundland to Minnesota, brought their tools for masonry with them, stayed in the Kensington area for several weeks –(the time that was needed to extract and work a sizable rectangular stone and to carefully execute a runic inscription), in a fashion unlike the way of living practiced anywhere in the world of Norse civilization.

We are also urged to believe that all this happened at a time when the runic script had fallen almost into oblivion (having been superseded by the Latin alphabet). In the second half of the 14th century, runes had become a matter of antiquarian interest, and the knowledge of it was preserved only in the minds of some learned literates (Elliott 1996: 335). Even the use of runes for wooden message sticks (called *rúnakefli* in Norwegian), with their long medieval tradition, had ceased by the time when the Kensington exploration allegedly took place. Common sense is likely to revolt against such a fanciful scenario.

There is another inconsistency in the "runic" inscription of the Kensington Runestone that casts a shadow on its claimed authenticity, and this is the stylistic technique of the textual composition. Its tenor is that of a narrative text in the style of the Icelandic sagas, which is atypical of rune stones. "The typical runic inscription from the Viking Age is a raised stone with a memorial inscription to a person who has died. It is not necessarily placed over the person's grave; such stones were also raised in public places or by bridges and roads so that they could be read by passers-by. . . . The inscription may also have an addendum about the circumstances of the death or some description of the dead person" (Spurkland 2001: 86).

The motivation for erecting rune stones was manifold, connecting the writer(s) of a text, in communicative reciprocity, with potential readers. The commemorative function of the inscription on the stone only made sense if there were people who would read this text, and the readers would honor, in their memory, both the deceased person and the donor of the rune stone. Consequently, rune stones were intended to be noticed—as landmarks—and their texts to be read by other people who were familiar with the runic script. According to these criteria, the Kensington Runestone does not fall within the category of Norse rune stones.

Why would Norse explorers erect a rune stone in the middle of nowhere, in the awareness that, after they had left the place, nobody would have been around who was capable of reading its runic text? Since there was no audience to receive and appreciate the contents of the runic inscription, there was no motivation whatsoever to make an effort to erect a rune stone.

If the Kensington exploration had occurred in reality, its participants, in all probability, would have chosen a different means of communication for their memory to become immortal. Those who would have survived the fights with the *skraeling* would have returned home and told the story of their spectacular expedition. And they would have certainly made mention of their labor to erect a rune stone on-site. The Europeans in their 14th-century crisis would have loved to have been cheered up by a heroic exploration story and distracted from their misery, and generations of storytellers would have embellished the narrative. But no such saga or story is known from Norse or Icelandic literature.

Documentary Evidence for Runic Inscriptions in Greenland

Only one runic inscription on a small stone is known from Greenland, the Kingiktorssuaq Stone. It was discovered by a Greenland Inuit on an island (located at approximately 73 degrees north) off the shore of west-central Greenland, in the Upernavik district, north of Disko Bay. Disko Bay is in the area of the northern hunting grounds (Nordsetur in Norse), which were frequented by the Norse colonists. The stone material is phyllite, a mineral, which is available in the area. This miniature rune stone records the names of three hunters and is dated to the second half of the 13th century. The text reads as follows: "Erlingur Sigvatsson, Bjarni Thordarson, and Enridi Odsson built Saturday before Rogation Day a cairn" (Gulløv 2000: 320). The authenticity of this rune stone has remained undisputed.

One could argue that the motivation of the inscription on this small stone resembles that of the recording on the Kensington Runestone. In both cases, the presence of Norsemen (of explorers or hunters, respectively) is documented. There is a significant difference, though. The writer of the inscription of the Kensington Runestone could not expect to have an audience of readers, whereas the writer of the runic inscription from Greenland was sure that other hunters who would visit the northern hunting grounds would read the text and learn to whom the cairn belonged.

This runic document is the last evidence for Norse presence in west-central Greenland. Since about 1300, the *skraeling*, the Greenland Inuit, became more and more hostile, engaged in clashes with the colonists, and gradually pushed the Norse down to the south. By the middle of the 14th century, the western settlement had already been abandoned as a result of constant attacks from the Inuit. The last traces of Norse community life in Greenland are lost at the beginning of the 15th century.

References and Further Reading

Anderson, Rasmus B. *America Not Discovered by Columbus: An Historical Sketch of the Discovery of America by the Norsemen in the Tenth Century*, 4th ed. Chicago: S.C. Griggs and Co., 1891.

Cohen, Robin, ed. *The Cambridge Survey of World Migration*. New York: Cambridge University Press, 1995.

Cox, Steven L. "A Norse Penny from Maine." In *Vikings: The North Atlantic Saga*. Edited by William W. Fitzhugh and Elizabeth I. Ward. Washington DC: Smithsonian Institution, 2000.

Daniels, Peter T., and William Bright, eds. *The World's Writing Systems*. New York: Oxford University Press, 1996.

Elliott, Ralph W. V. *Runes: An Introduction*. New York: St. Martins Press, 1996.

Gjerde, Jon. "The Scandinavian Migration." In *The Cambridge Survey of World Migration*. Edited by Robin Cohen. New York: Cambridge University Press, 1995.

Gulløv, Hans-Christian. "Natives and Norse in Greenland." In *Vikings—The North Atlantic Saga*. Edited by William W. Fitzhugh and Elisabeth I. Ward. Washington, DC: Smithsonian Institution Press, 2000.

Houston, Stephen D., ed. *The First Writing. Script Invention as History and Process*. New York: Cambridge University Press, 2004.

Knirk, James E. "Runes and Runic Inscriptions." In *Medieval Scandinavian: An Encyclopedia*. Edited by Phillip Pulsiano and Kirsten Wolf. New York: Garland, 1993.

Lambropoulos, Vassilis. *The Rise of Eurocentrism. Anatomy of Interpretation*. Princeton, NJ: Princeton University Press, 1993.

Lynnerup, Niels. "The Greenland Norse. A Biological-Anthropological Study." *Meddelelser om Grønland: Man and Society* 24 (1998): 1–149.

McGovern, Thomas H. "The Demise of Norse Greenland." In *Vikings: The North Atlantic Saga*. Edited by William W. Fitzhugh and Elizabeth I. Ward. Washington, DC: Smithsonian Institution, 2000.

Pulsiano, Phillip, ed. *Medieval Scandinavia—An Encyclopedia*. New York: Garland Publishing, 1993.

Sawyer, Birgit. *The Viking-Age Rune-Stones. Custom and Commemoration in Early Medieval Scandinavia*. New York: Oxford University Press, 2000.

Schoolcraft, Henry R. *Information Respecting the History, Condition, and the Prospect of the Indian Tribes of the United States Collected and Prepared Under the Direction of the Bureau of Indian Affairs per Act of Congress of March 3rd, 1847*, 4 vols. Philadelphia: Lippincott, 1853–57.

Smith, Kevin P. "Jasper Cores from L'Anse aux Meadows." In *Vikings: The North Atlantic Saga*. Edited by William W. Fitzhugh and Elizabeth I. Ward. Washington, DC: Smithsonian Institution, 2000.

Spurkland, Terje. *Norwegian Runes and Runic Inscriptions*. Rochester, NY: Boydell Press, 2001.

Squier, Ephraim G., and E. Hiram Davis. *Ancient Monuments of the Mississippi Valley: Comprising the Results of Extensive Original Surveys and Explorations*. Washington, DC: Smithsonian Institution, 1848.

Wallace, Birgitte L. "L'Anse aux Meadows." In *Medieval Scandinavian: An Encyclopedia*. Edited by Phillip Pulsiano and Kirsten Wolf. New York: Garland, 1993.

Wallace, Birgitte L. and William W. Fitzhugh. "Stumbles and Pitfalls in the Search for Viking America." In *Vikings: The North Atlantic Saga*. Edited by William W. Fitzhugh and Elizabeth I. Ward. Washington, DC: Smithsonian Institution, 2000.

Williams, Henrik. "Reasons for Runes." In *The First Writing: Script Invention as History and Process*. Edited by Stephen D. Houston. New York: Cambridge University Press, 2004.

Wolter, Scott F. and Richard Nielsen. *The Kensington Runestone: Compelling New Evidence*. Minneapolis: Lake Superior Agate Publishing. 2005.

2

The Ancestral Puebloans lined up their communities so that, although miles apart, they could signal each other with fires by line of sight to communicate.

PRO	Linda Karen Miller
CON	Peter N. Jones

PRO

Over a millennium ago, a number of imposing masonry pueblos, or great houses, were erected in Chaco Canyon in northwestern New Mexico. These great houses, such as Pueblo Bonito, are some of the best-preserved standing ruins in the American Southwest. The Anasazi (now called the Ancestral Puebloan) constructed these settlements at Chaco Canyon in the San Juan River Basin of northwestern New Mexico, later at Mesa Verde in southwestern Colorado, at Canyon de Chelly in northeastern Arizona, and at many other places by the end of the first millennium CE.

The vast landscape that was once the territory of the Anasazi encompasses the Colorado Plateau but extends from central New Mexico on the east to southern Nevada in the west. The northern boundary winds through southern Nevada, Utah, and Colorado as the southern boundary follows the Colorado and Little Colorado rivers in Arizona and the Rio Puerco to the Rio Grande in New Mexico, with extensions eastward toward the Great Plains on the Cimarron and Pecos rivers and the Galisteo Basin. Archaeologists divide the Anasazi area into subregions named for geographic regions with certain ceramic technology: from west to east, there are the Virgin, Kayenta, Tusayan, De Chelly, and Winslow, Northern San Juan, Cibola (or Chaco), and Rio Grande.

The Anasazi people succeeded in developing what is called the Chaco Phenomenon between 900 and 1020 CE in planned stages to a time of greatness between 1020 and 1120. At their cultural zenith during the first third of the second millennium, the Anasazi emerged as the preeminent community planners, architects, and builders of the plateau cultures. The remnants of the time of greatness remain in the ruins of immense ruined towns, villages, earthen mounds, roads, and irrigation works.

John Macrea Washington led a military expedition into the Navajo country in 1849. His surveyor, Simpson, wrote descriptive notes about the ruins, and an

---

The Ancestral Puebloans

The ancient Pueblo people, sometimes called by the Navajo name *Anasazi*, probably emerged around the 12th century BCE and descended from small bands of hunter-gatherers. They're best known for their homes—pit houses originally, which developed into the more elaborate and inaccessible dwellings—and pottery, which in later periods was often highly decorated. Like the Sumerians and some other ancient civilizations that displaced earlier settlements, it isn't clear from where they migrated. Much of the more sophisticated development of the Anasazi occurred in the temperate, fertile period of the 10th and 11th centuries CE. Like other North American cultures, they suffered from the run of droughts and cold summers that lasted from about the middle of the 12th century to shortly before European contact.

---

artist made sketches. The resulting manuscript was first published in 1850. An archaeological survey was conducted years later. During the 1930s, nearly 10 miles of ruins were excavated. The project was completed in 1936, with a reported 200 sites mapped and surveyed. The first history of research in Chaco Canyon was written by Donald D. Brand in 1937.

The Anasazi originally lived in pit houses, semisubterranean dwellings with roofs made of natural materials. Eventually the Anasazi moved from the pit houses to adobe houses built above ground and often in large rock shelters in cliffs. Chaco contains 16 great houses (two are unfinished). It contains over 60 kivas (religious rooms). It is four stories tall and covers nearly two acres of land.

Archaeologically, a site consisted of a single "focal" site of 75 to 400 rooms, with 2 to 13 satellite sites of 2 to 25 rooms each. In addition to being large, each of the focal sites is distinguished by the presence of an open plaza area and a "spinal" room block and by being positioned along the back edge of the site. All of the focal sites are situated in defensible positions—hilltops, knolls, or isolated small mesas overlooking the land around the satellite sites in the cluster. Significantly, the focal sites are located so that there is a clear line of sight with each of the others and overlooking all the primary access routes leading into the valley. Above the cluster at valley level, the strategic position of the focal sites point to some degree of valley-wide cooperation and coordination. In their defensive locations, the focal sites are all carefully positioned so as to collectively oversee all access routes into the valley and at the same time maintain visual contact and communication between communities.

The importance of establishing visual contact between focal sites is reinforced by the discovery that a deep notch was excavated prehistorically in a hill slope that otherwise would have interfered with the line of sight from one site to the other. The notch then allowed visual contact and communication between the two focal communities. Jonathan Haas (1990) discusses that the visual network of defensive sites extends beyond the Long House Valley in Arizona to the east. It also

indicated that by 1250 there was a pan-valley, supravillage interaction and net-work in the valley that corresponded to a tribal form of organization.

Questions about the Great Houses

Many aspects of the great houses baffle researchers. How did the people have the manpower to build the massive structures? Were the roads part of a closely integrated communication system? Chaco experts have suggested that tower kivas were facilities in elevated positions that were used for communication between these great houses and the Chaco core. At several of the great houses, investigators have discovered these cylindrical masonry towers—"tower kivas," which are defined as multistory kiva-like structures that rose for two or three stories above the rest of the great house structure.

The towers may have served, according to speculation, as ceremonial structures, communication towers, lookout towers, or rudimentary lunar and solar observato-ries. The Chaco World Database revealed that 9 of 10 tower kivas in its archives are about 25 miles apart from one another, perhaps close enough for signaling. However, another study of tower kivas at the southern great houses of Kin Ya'a and Haystack showed that even if intermediate stations were built between them, they were not built to communicate with one another. Some theorize that Chaco Canyon was a central place, the ceremonial nexus for the outlying communities, and that the roads were basically ceremonial symbols of the great houses. Dozens of small towns lay along the roads, 60 miles to the north and 60 miles to the south.

Archaeologists rarely looked beyond the areas immediately surrounding their excavations. However, according to Stephen Lekson, a study over a dec-ade ago proposed that:

> three of the region's largest and most important ancient centers were linked by a 450–mile meridian line—Chaco Canyon, in New Mexico; Aztec Ruins, 55 miles due north near the Colorado state line; and Casas Grandes, 390 miles due south in Chihuahua, Mexico. Chaco and Aztec were also connected along the meridian by a road known today as the Great North Road.
>
> Tree-ring dates indicate that Chaco was occupied mainly from A.D. 850 to 1125 (after which its population was greatly diminished), Aztec from 1110 to 1275, and Casas Grandes from 1250 to 1500. As Chaco was dying out a new center was being created on the same meridian at Aztec. In commemoration of the earlier site, the new one was connected to it by the Great North Road. When Aztec collapsed and the Four Corners area was abandoned, Casas Grandes arose due south of Chaco. (Lekson 1997)

These centers were sequential from 900 to 1450 CE. The most unique feature about their positional alignment is that the capitals were all built on the same meridian.

Chaco was the center of the 11th-century Pueblo world. Chaco was very likely a center of some kind, very possibly a sacred meeting place, where the tribes would gather on special occasions for ceremonies or to renew old alliances and to transmit the ancient stories and songs after being notified by the signal fires. The Chacoans might have used the line-of-sight visual communication network of pinnacle shrines to communicate to the people to come to the center from across the Colorado Plateau for ceremonies or to warn them of impending trouble. The people came in massive numbers, moving in procession across the roads.

Ancestral Puebloans and Sight Lines

Many scientists believe that the Chacoans developed a communication system. Certain towers and mesa or ridge top ruins seem likely placed for receiving and transmitting line-of-sight messages by smoke signals during the day and interrupted fire-light signals at night. The canyon is ringed with four sites that are on high points with exceptional visibility. By the end of their occupation, the sites had a functional defensive configuration, and they were all visible to one another. This line-of-sight system, presumably operated with smoke or mirrors, paralleled the more famous Chacoan roads, and may prove to be more important to the understanding of Chaco and its world.

Questions about why and how populations congregate are important ones in archaeology, anthropology, and geography. The alignment of these centers and the roads is the topic of this debate. Does the alignment reflect religious connections or communication connections? Why did these populations congregate in such a way? I contend that the alignment was for communication purposes.

Since the 1970s, scholars have known about the existence of an elaborate line-of-sight system spanning large portions of the Chaco region. Archaeological surveys done in Chaco Canyon by Alden Hayes and Tom Windes (1975) revealed evidence of the communication link. It was their excavation work during the late spring of 1973 of Site 29 on the mesa above Penasco Blanco that discovered the line-of-sight network that extended from the buildings of Chaco Canyon to the surrounding region and perhaps beyond to its very farthest corners that could not have been accidental but were carefully selected.

Three shrines, with five miles between the outermost, are all in sight of each other, and among them command a view of all the major pueblos in the canyon as well as Pueblo Alto to the north and Kin Ya'a, 27 miles to the south. Hayes and Windes discovered that all of the Bonito phase sites in the canyon and on both sides of it are visible from one or more of the three shrines, with the exception of the pueblo of Una Vida. Una Vida is located against the canyon wall under the Great Kiva, Kin Nahasbas. From Kin Nahasbas one can look directly down upon the tower kiva in the Vida's west wing. All three shrines are in view of Tsin Kletzin. From Tsin Kletzin, one can see the smaller pueblo and tower

Pueblo Bonito, in New Mexico's Chaco Canyon, was the center of Ancestral Puebloan culture and commerce between approximately 900 and 1150 CE. (Clive Ruggles)

kiva of Kin Klizhin, 6.5 miles west. Kin Bineola, 10.5 miles west, is invisible, but a small tower-like structure looks down on Kin Bineola and can be seen from Tsin Kletzin.

Thus they discovered that all of the contemporary large pueblos of the Chaco were locked into a network, with the exception of Pueblo Pintado. Because these structures were placed at the few limited points and within the narrow restricted areas that permitted intervisibility, Hayes and Windes were led to believe that they were relay points for signals. Huerfano Mountain, in turn, has a direct line of sight to Pueblo Alto. Messages could have been passed from Chimney Rock at the northeastern edge of the Chacoan region to downtown Chaco in a matter of minutes if people were manning the stations at Huerfano Mountain. The signaling system would have required careful management, staffing, and coordination. Shrines were often combined with signal stations. Hayes and Windes found that the shrines were linked but that there was no evidence that the shrines had fires. They also found few actual tower kivas. These observations were deduced from plotting the contour lines on U.S. Geological Survey topographic sheets. They could also be verified with the eye and from the ground. The evidence from Chaco Basin has been more commonly preserved and studied.

In the spring of 1975, further investigation found another one-acre tract on Chacra Mesa from which the Pueblo Pintado, Kin Ya'a, and Tsin Kletzin could be seen. The Charcra station did not command a view down into the canyon but was visible from a fifth arc, located at the only point on the southeastern edge of

South Mesa from which Kin Ya'a could be seen to the south and Una Vida and Kin Nahasbas to the north. Still another shrine, a half mile west, revealed a view of the Chacra point that included Kin Ya'a and Kin Klizhin to the west and the mesa immediately above Kin Bineola. Consequently, the seventh wall arc on the southwestern edge of South Mesa was in a position to communicate with those same three pueblos but not with the Chacra shrine, nor with anything in the vicinity of the canyon. These researchers concluded that it would be possible to link the entire population within an area about 30 miles square with three to four fires or smokes, in spite of the broken terrain. If signals were the function of these sites—and they believe that they were—one would expect to find a similar one on Huerfano Butte, about 30 miles north of Chaco, which is visible from several spots on the San Juan River, from Aztec, and even from Mesa Verde.

An experiment by members of the Chaco Center, research staff tested the line-of-sight connections among these features with signal flares to suggest that people intentionally positioned these features to ensure intervisibility. The archaeologists equipped themselves with railroad flares and stationed themselves at certain points. At 9 P.M. the flares were lit and visual communication was established between each signaling station and its two neighboring stations. Every station could observe the lights at one or more of the other stations.

Site location based on line-of-sight communication has been well documented for the Kayenta area by Haas and Creamer (1993). They made the observation that line-of-sight communication linked sites within a cluster, but there were no visual links between the clusters. The people of Klethla Valley removed part of an intervening cliff to create a visual link between two sites. There is also evidence at El Morro of a similar pattern of site intervisibility at the Hole-in-the-Rock site, where it's possible a hole was cut into a mesa remnant to provide a view of other sites. There is also evidence in the Pecos area, the Verde Valley, and New River area. Within many of the clusters of pueblos, there appear to have been small pueblos located on high points so as to provide a visual link between large pueblos located in low-lying areas.

While most of these small pueblos have never been examined to determine whether they do provide such a link, they were present in the Chama, Pecos, Petrified Forest, Chacra Mesa, Grasshopper, and the Upper Little Colorado River settlement clusters. In some cases, large pueblos were placed in high places, even though other pueblos of the same size in the same cluster were situated in low places. This suggests the large pueblos situated on high places were there not simply for defense, but also to facilitate line-of-sight linkages between pueblos with their cluster. Clusters with such large, high sites include Zuni, Newton-Rattail, Acoma, and Anderson Mesa.

The argument that the purpose of the line-of-sight settlement locations was for communication rather than surveillance can be found in the distance between sites. The locations of sites in El Morro Valley and in Kayenta meant

that they could have seen each other. As they were three miles apart, it would not have been possible to observe details of activities taking place at another pueblo or even to see who was coming and going. Thus it seems that the pueblos were situated not to "spy" on potential enemies but were located to enable signaling for mutual support. The presence of towers on sites does not automatically mean they were used for line-of-sight communication. Many of these towers may have served as watchtowers.

The number of documented cases for line-of-sight communication is small, because this is a badly neglected area of investigation. It seems though, that intervisibility could have been a common feature of the late period settlement landscape. Testing for such visual links between sites is relatively easy and would be a productive activity for researchers.

Historical Interpretations

According to Steve LeBlanc (1997), site intervisibility was closely related to the presence of site clusters. If sites were located to be mutually supporting, then they must have been able to send for help during an attack. While there were other means of asking for help, there were certainly site clusters that did not have intervisibility in some of the cliff dwellings. Signaling would have been one of the best and fastest ways to communicate. Site intervisibility has rarely been investigated, although there are some good examples from the Southwest. In spite of these considerations, intervisibility can be determined with minimal excavation, and its value is high for showing evidence of warfare and for defining boundaries of site-cluster alliances. The intervisibility can be demonstrated, and the ethnographic explanation for such communication is defense. There is also debate among scholars as to the distance that signaling is practical. Signals can be very simple if mutual support is needed. Signals that work only at night would have their limitations, and sunlight-reflecting mirrors would not be useful in all circumstances. Some combination of fire at night and smoke during the day would have been both practical and adequate for many parts of the world.

Some scholars have suggested that line-of-sight links defined alliance boundaries. It is possible that clusters of sites belonging to the same alliance would have contained sites that were intervisible. It was significant that allied sites were not intervisible with enemy clusters. The lack of intervisibility between sites probably defined the alliance boundary if sites within the boundary were intervisible. If a number of sites within the boundary were intervisible, and a gap is found where there is no intervisibility with another site cluster, it's probably evidence for an alliance boundary.

To address the question of kiva intervisibility, researchers conducted a recent GIS–enhanced study to determine whether two of the tower kivas could have signaled each other using intermediate relay points, according to a study

by John Kantner (2009). A computerized model of the now-collapsed tower kivas at Kin Ya'a and Haystack and the surrounding landscape convinced researchers that even if intermediate relaying stations existed, the tower kivas were not likely meant to communicate with each other. The computer did identify potential relay points between the tower kivas, but no archaeological features were found at these stations. These points were already visible from the ground, which meant that the towers were superfluous and did not achieve long-distance visibility. However, their extra height did greatly improve visibility of the immediate community area. This interplay between the visible and the invisible emphasizes Chaco as the center place.

It is not clear what methods of communication would have been used between these intervisible pueblos, but fire and smoke are likely candidates. The other possibility for communication is mineral mirrors. James Riddle (2005) believes that the Chacoans may have used a primitive source of mirrors. So he tested several mineral specimens. The best results were from mica, which could be use to create a larger field of reflection as it diffused the light. In tests using mica as a reflective material, Riddle was able to send a signal 3.92 miles, which would certainly be far enough to communicate with the next settlement. He believes that a set of signals, corresponding to a number of flashes, could have been worked out between the settlements. Such an assertion, however, is difficult to verify with certainty, as mica is quite fragile, and such mirrors made of the mineral would not be likely to survive today. However, mica was a readily available reflective material that the Chacoans certainly mined. The tests he performed proved that thin mica slabs could be used possibly at distances of 10 or more miles, depending on the size of the slab, and even greater distances if the sun is low in the sky. Mirrors would have been treasured and might have been destroyed rather than have them fall into enemy hands. A good description of the Chaco Canyon settlement layout would be a central place, ringed by forts which would signal each other for help. The signaling system was known to extend far outside of Chaco Canyon, such as for the Sonora and the Rio Grande areas (Riddle 2005).

Road Networks

An extensive road network connected the Chacoan economic, political, and religious systems. The paths connected the great houses and great kivas in the canyon to distant, more outlying villages and sacred sites. People may have used these kivas' elevated positions, combined with still-undiscovered "repeater" stations, to relay signals from community to community. Linda Cordell (1994) noted that this was the greatest aboriginal land communication system north of Mexico. Researchers have discovered 400 miles of roads in northwestern New Mexico. The roads from Chaco Canyon stretch 40 miles to the San Juan River. The roads are straight from point to point and skip major obstacles.

Horizontal directions are inscribed on the Chacoan landscape, suggesting that the Chacoans thought of the canyon as the center place at the intersection of both vertical and horizontal planes. Chaco Canyon sits at the geographic heart of the San Juan Basin. This centrality may be one reason why Chaco became such an important location. There are also four horizontal directions expressed through roads and through building alignments and orientations. The function of the roads has been debated. The alignments might not been primarily for transportation but for a symbolic package of architectural features that included great houses, great kivas, and earthworks.

To the Anasazi the straight road signified the spiritual path, the means of direct return to the heart, the way of life, just as the rays of the sun always take a straight course. These "roads," or cleared linear alignments, have long been recognized across the Chacoan world, although their function has been debated. In the 1980s, researchers believed that the roads extended out from Chaco like spokes on a wheel. Later ground verification by the Bureau of Land Management demonstrated that the road network was not that extensive. The alignments primarily connect Chaco with other resources and may only extend a short distance from the great house before disappearing. The alignments entering Chaco help the pedestrian to maximize the juxtaposition of the visible and invisible. The two longest roads—the North Road and the South Road—fix Chaco as a center place balanced between low and high places. Both roads extend for more than 30 miles each from Chaco. The east-west roads were less defined.

The remarkable network of fire-signal or mirror-signal stations paralleled the straight roads. Researchers believed that a relationship existed between the line-of-sight signal stations and the configuration of the prehistoric road system. Through the roads, fundamental spiritual ideas were ceremonially reenacted physically on a periodic basis, instilling the wisdom of the ancestors into every member of the tribe from the earliest years.

Solar and Lunar Cosmology of Major Buildings

Work done by the Solstice Project showed that Chacoan construction techniques incorporated elements of solar and lunar cosmology in the buildings' internal layout, directional orientation, and in the geographic layout of the settlements. In her 1997 article, "The Primary Architecture of the Chacoan Culture: A Cosmological Expression," Anna Sofaer stated:

> Twelve of the fourteen major Chacoan buildings are oriented to the midpoints and extremes of the solar and lunar cycles. The eleven rectangular major Chacoan buildings have internal geometry that corresponds with the relationship of the solar and lunar cycles. Most of the major buildings also appear to be organized in a solar-and-lunar regional pattern that is symmetrically ordered about Chaco Canyon's central complex of large ceremonial buildings.

These findings suggest a cosmological purpose motivating and directing the construction and the orientation, internal geometry, and interrelationships of the primary Chacoan architecture. (Sofaer 1997: 90)

Therefore, astronomical orientation was a vital part of Chacoan construction techniques. Some sites are critically situated for multiple views.

Precise Line of Sight Calculation

How could these primitive people have calculated the line of sight so carefully? William Calvin (1997), from the University of Washington, in his article on surveying north-south lines without modern instruments, proposes that "surveying long lines that are not north-south, such as Chaco's other ancient roads, can use a two-observer method: build large signal fires at each end of the desired path on a still day and then maneuver in between." He suggested that "an experienced team of 'utility pole' haulers" could raise the pole and steady it with guylines. Then, they would follow the pole's shadow over the course of the day using a rope from the base of the pole. Next, he says, a second team would be sent south with another pole, with which they would determine the longest distance over which a sight line could be seen, usually on a high point. Then, at the first pole, a surveyor looked to determine whether the pole or smoke from a signal fire was aligned, telling the team with the second pole to relocate if necessary. After the poles were properly aligned, the initial team moved their pole to a further location, and then the second repeated the process, leapfrogging each other to create a long north-south survey line. By using signal fires (and their smoke plumes) rather than poles, direct sight lines might not even be necessary to have a workable system.

Conclusion

The above examples have been given to show that there is evidence of an Anasazi civilization communication system among the various Rio Grande pueblos. Researchers still do not agree on how this communication was made or how far it reached.

One question often asked of the Chacoan regional system is how this could possibly work. How could the Chaco affect and control a great house 240 km (eight days' walk) distant? The answer has to be communication. The clearness of Southwestern skies, its open landscape, broad vistas, and a large complex line of sight communications system spanned large portions of the Chaco region. Studies have shown that the line-of-sight system extends over much of the northern San Juan Basin.

Recently there was a discovery in Peru of a tower that researchers believe was a solar observatory. The buildings there were also very well aligned as were those in Chaco. The ancestors of the Incas, who probably were contemporaries of the

Anasazi, did this as well. While visiting Korea, I saw another solar observatory that was built around 622. Could all of these be connected? Could this be the answer for the alignments in all areas: Were they all communicating with the sun to tell the seasons? Further research will need to be done.

By the 14th century, the Anasazi abandoned the Chacoan region, the Mesa Verde region, and the southwestern Colorado Plateau, leaving behind empty cities of stone. Did their line-of-sight communication fail them, and were they invaded by the more advanced Aztec? It is the pueblos of northwestern Arizona and western New Mexico that greeted the Spanish expeditions into the Southwest in the 16th century. The last of those pueblos now stand as living monuments to the endurance of the Anasazi traditions, the cultural strength, the adaptability, and the resourcefulness of an ancient people. Chaco stands as a place where people achieved monumental things. These achievements were accomplished by this intricate line-of-sight communication, which was far ahead of its time.

References and Further Reading

Calvin, William H. "Leapfrogging Gnomons: A Method for Surveying a Very Long North-South Line without Modern Instruments." 1997. Available at http://williamcalvin.com/1990s/1997gnomon.htm (accessed January 8, 2010).

Cordell, Linda. *Ancient Pueblo Peoples*. Montreal: St. Remy Press, 1994.

Cordell, Linda. *Prehistory of the Southwest*. Orlando, FL: Academic Press, 1984.

Fagan, Brian. *Chaco Canyon: Archaeologists Explore the Lives of an Ancient Society*. New York: Oxford University Press, 2005.

Gabriel, Kathryn. *Roads to Center Place*. Boulder City, NV: Johnson Books, 1991.

Hass, Jonathan. *The Anthropology of War*. New York: Cambridge University Press, 1990.

Hass, Jonathan and Winifred Creamer. *Stress and Warfare Among the Kayenta Anasazi of the Thirteenth Century A.D*. Chicago: Field Museum of Natural History, 1993.

Hayes, Alden C., and Thomas Windes. "An Anasazi Shrine in Chaco Canyon." In *Papers of the Archaeological Society*. Edited by Theodore Frisbie. Norman, OK: Hooper Publishing, 1975.

Hayes, Alden C., David Brugge, and James Judge. *Archeological Surveys of Chaco Canyon New Mexico*. Washington, DC: National Park Service, U.S. Department of the Interior, 1981.

Kantner, John. *The Ancient Puebloans from A.D. 900 to A.D. 1150*. Santa Fe: School for Advanced Research, 2009. Available online at http://www.sarweb.org/sipapu/history/timeline/timeline900.html. (accessed August 5, 2010).

LeBlanc, Steve. *Prehistoric Warfare in the American Southwest,* Salt Lake City: University of Utah Press, 1997.

Lekson, Stephen. *The Archaeology of Chaco Canyon: An Eleventh-Century Pueblo Regional Center*. Santa Fe, NM: School of American Research Press, 2006.

Lekson, Stephen. *The Chaco Meridian*. Walnut Creek, CA: Altamira, 1999.

Lekson, Stephen. "Rewriting Southwestern Prehistory." *Archaeology* 50:1 (January/February 1997), abstract. Available online at http://www.archaeology.org/9701/abstracts/southwest.html (accessed August 5, 2010).

Noble, David, ed. *In Search of Chaco*. Santa Fe, NM: School of American Research Press, 2004.

Reed, Paul F. *The Puebloan Society of Chaco Canyon*. Westport, CT: Greenwood, 2004.

Riddle, James. "Prehistoric Heliographs." July 21, 2005. http://myweb.cableone.net/lkd7aoilthe.htm (accessed January 8, 2010).

Rohn, Arthur, and William Ferguson. *Pueblo Ruins of the Southwest*. Albuquerque: University of New Mexico Press, 2006.

Sofaer, Anna. "The Primary Architecture of the Chacoan Culture: A Cosmological Expression." In *Anasazi Architecture and American Design*. Edited by Baker H. Morrow and V. B. Price. Albuquerque: University of New Mexico Press, 1997.

Windes, Thomas. *Investigations at the Pueblo Alto Complex Chaco Canyon, New Mexico*, Vol. I. Santa Fe, NM: National Park Service, U.S. Department of the Interior, 1987.

CON

The complex, sophisticated, and extensive ruins encountered throughout the four corners region of the American southwest have along inspired scholars. James Simpson of the U.S. army, one of the first Euro-Americans to describe the region when he visited the area in pursuit of Navajos in 1849, concluded that the superior construction of the ruins allied them with the Aztec, who, legend says, moved from the north to the south into Mexico. Professionally trained archaeologists working in the area during the 1920s and 1930s were skeptical of the Aztec connection, however, as it seemed as if such a great distance between the two regions would precluded any cultural connection. Edgar Hewett (1936), in *The Chaco Canyon and Its Monuments*, and Neil Judd (1954), in "The Material Culture of Pueblo Bonito," for example, disagreed with the Aztec cultural connection hypothesis and argued rather that the builders of the ruins were the ancestors of the contemporary Pueblo peoples. This theory was hard for many to believe because

the ruins of the four corners region were more extensive and architecturally grander than those of the contemporary Pueblo people. Doubters of the local Puebloan connection countered that it was unreasonable to assume that the Puebloan peoples would use a less-complex architectural building pattern if they had previously been such architectural masters.

As a result, between 1950 and 1970 a return to the Mesoamerican connection was a common theme among scholars working in the Southwest, including Edwin Ferdon (1955), Charles DiPeso (1974), J. Charles and Ellen Kelley (1975), and Robert Lister (1978). Rather than Aztecs, however, they suggested that a small group of Toltec traders could have organized the prehistoric Ancestral Puebloan into a period of cultural florescence that centered on turquoise, which was highly valued among the societies in Mesoamerica.

As part of this resurgence in a Mesoamerican connection, scholars attempted to find specific cultural correlations between Mesoamerican societies and the Ancestral Puebloan. For example, scholars noted that, just like in Mesoamerica, many of the Ancestral Puebloan communities and buildings appeared to be aligned with one another or with various geophysical phenomena. Because many of the Mesoamerican societies aligned their buildings for long-distance signaling purposes, a similar analogy was transplanted onto the Ancestral Puebloan. Like the pyramids of Mesoamerica, scholars initially argued that the communities and buildings of the Ancestral Puebloan were aligned so that they could signal one another with fires by line of sight.

The recently discovered time depth of the Ancestral Puebloan development, beginning around 700 CE, coupled with the late date of demonstrably "Mesoamerican" objects in the area (macaws, copper bells, etc.), has cast doubt on any Mesoamerican connection concerning the alignment of Ancestral Puebloan communities and buildings, but it did not necessarily rule out that the Ancestral Puebloan used some form of large-scale signaling system. For example, Thomas Windes (1987), in "Population Estimates," argued that the Ancestral Puebloan used a line-of-sight signaling network. However, this hypothesis can no longer be held favorable, because there is now a fairly large body of empirical evidence indicating that the Ancestral Puebloan aligned their communities and buildings for spiritual reasons and not for signaling purposes.

Support for Spiritual Alignment Theory

Early support for the spiritual alignment of Ancestral Puebloan buildings came from the archaeoastronomic work of John Carlson and J. James Judge (1987) and of Ray Williamson (1984). Presently architectural features found throughout the Ancestral Puebloan region, the extensive network of roads linking communities to one another across the area, and the extensive and complex archaeoastronomic features involving the specific layout of buildings, pictographs, and geophysical

phenomena all argue that the alignment of Ancestral Puebloan communities was spiritual.

The Ancestral Puebloan were a people who occupied the Colorado Plateau and surrounding areas of the southwestern United States, roughly bounded by the Rocky Mountains to the north and the Mogollon Rim to the south. The Ancestral Puebloan's cultural origins began around 700 CE and lasted through the 1200s, with an apex occurring between 1020 and 1100 CE. During this period, these people built numerous communities throughout the region that included large-scale architectural structures, extensive road networks, and modified landscape features and earthworks. The Ancestral Puebloan were primarily farmers who grew a variety of crops in lowland basins of the four corners region of the southwest, and there is evidence of long-distance trade and exchange with other cultural groups. Most of these communities contained between 10 and 100 households, were oriented in a southeasterly direction, and contained a number of characteristic architectural features. Two of the most common architectural features that have come to be associated with the Ancestral Puebloan are large-scale, multistoried buildings known as great houses, and circular or enclosed structures known as grand kivas.

The most spectacular of the Ancestral Puebloan communities was constructed in Chaco Canyon, located in northern New Mexico. Recent archaeological evidence suggests that Chaco Canyon was one of several large centers that encompassed more than 200 communities spread over 110,000 square km of the northern southwest and that these communities and structures within each community were purposefully aligned for spiritual reasons. As mentioned, the roots of Ancestral Puebloan culture have origins in the earlier Basketmaker III and Pueblo I (775–825 CE) time periods, but by 850–900 CE, when the first great houses were built, such as the early houses at Una Vida and Pueblo Bonito, evidence of purposeful construction and alignment of buildings across the Ancestral Puebloan region are encountered. These two great houses demonstrate, in fact, that the Ancestral Puebloan purposefully aligned their buildings at a very early date and that this alignment was for spiritual purposes, as these two great houses are aligned with prominent landmarks on the adjacent buttes, Fajada and Huerfano, respectively.

The Solstice Project

Organized in 1978 by artist and archaeoastronomer Anna Sofaer, the Solstice Project conducts research on and makes films about the Sun Dagger, a spiral petroglyph used as a calendar by the Pueblo people. Archaeoastronomy is the study of astronomy as known to, and studied by, ancient peoples—encompassing everything from their recordings of celestial events to the theoretical use of constructions like Stonehenge and the Pyramids in measuring time and the positions of the stars. The interdisciplinary field is often highly speculative, as it is difficult to definitely prove the validity of a claim.

As discussed by David Doyel (2001) in *Anasazi Regional Organization and the Chaco System*, these attributes, along with others such as visibility within and between the communities and across the landscape, association with a great kiva, and stone masonry construction indicate that the Ancestral Puebloan spiritual landscape played an early role in architectural design and the location of each community. In fact, in "Recent Research on Chaco Prehistory," James Judge and J. Schelberg (1984) have convincingly argued that the great houses found within the Ancestral Puebloan region functioned primarily as nondomestic facilities that integrated the populations of geographically widespread communities, perhaps for specific events, such as community rituals or other spiritual observations.

Monumental Architecture

Incorporating this spiritual landscape across the Ancestral Puebloan region was an extensive series of roads and pathways stretching over 2,400 km (1,500 miles). These achievements were so grand compared to earlier cultural manifestations found in the region that many of the architectural features, such as great houses and grand kivas, have come to be known as monumental architecture. In Chaco Canyon, where the development of monumental architecture reached an apex, it is now agreed that the monumental architecture was constructed to convey, reinforce, and codify ideas about Ancestral Puebloan spiritual and cosmological order. This spiritually and cosmologically ordered landscape developed over time as the alignment and construction of architectural and earthen features progressed. Archaeological evidence indicates that by 1000–1040 CE, Ancestral Puebloan architecture expressed the basic tenets of a unified worldview: directionality, balanced dualism, and places of central importance. At the beginning of 1100 CE, however, evidence indicates that confidence in this newly formed ritual order was shaken by specific environmental and social developments as a result of population pressures, drought, and resource constraints. As a result, it has been hypothesized that leaders sought to reformalize the Ancestral Puebloan worldview by instituting a new building scheme across the region, which was centered at Chaco Canyon. Six new great houses were positioned on the landscape in a patterned, nested series of oppositional relationships. Furthermore, this reformalization of the Ancestral Puebloan landscape was legitimated, as discussed by Ruth Van Dyke (2004) in "Memory, Meaning, and Masonry," through direct alignments and indirect architectural references to the already existing architectural features, as well as, according to David Doyel and Stephen Lekson (2001) in "Regional Organization in the American Southwest," to various geophysical and astronomical phenomena.

As noted, a number of attributes, including formality, dimensional uniformity, and large-scale construction over a brief period of time, suggest that a vision, or

plan of the structure, existed before construction began. Great houses such as Pueblo Bonito and Chetro Ketl are among the best-preserved prehistoric pueblo ruins in the Ancestral Puebloan region that exhibit these attributes, with standing walls in excess of eight meters in height and several dozen rooms in each great house. These planned, monumental architectural structures represent a substantial investment of both labor and design. For example, although three centuries may have elapsed from initiation to completion at Pueblo Bonito, the final form of the building, according to Lekson (1984) in "Standing Architecture at Chaco Canyon and the Interpretation of Local and Regional Organization," was attained in essentially seven epic construction events. The timing and architectural characteristics of the major construction cycles, in fact, are reflected, according to Andrew Fowler and John Stein (2001) in "The Anasazi Great House in Space, Time, and Paradigm," in the construction histories of great houses throughout the Ancestral Puebloan region.

Dating the Construction

Archaeological evidence indicates that great house construction was initiated during the late 9th century and escalated through the early 12th century CE, during which time great houses and associated features appeared across the greater Ancestral Puebloan region, often in the midst of communities of small, domestic sites. As noted above, great houses are large, massive, usually multistoried structures built in planned construction episodes. Core-and-veneer architecture is typical of great house construction, and a number of veneers, or facing styles, have been recognized. Great houses are differentiated from smaller, domestic structures by their larger rooms and higher roofs, as well as by their overall size and scale. Beyond their planned design and construction history, other evidence supporting the spiritual alignment argument is the fact that most great houses contain enclosed kivas, or circular rooms, constructed inside rectangular rooms. Likewise all great houses are constructed to face the south/southeast, a pattern common to many Ancestral Puebloan sites, reflecting the importance of the sun and the cardinal directions in the Ancestral Puebloan spiritual landscape. Furthermore, great houses often are associated with one or more great kivas—large, semisubterranean circular structures usually considered to represent communal, public, integrative architecture. These attributes demonstrate that the great houses were purposefully aligned and positioned on the landscape and that this alignment was for spiritual purposes rather than for the purposes of signaling. In fact, Fowler and Stein have argued that the great house was a manifestation of an Ancestral Puebloan sacred technology in that the monumental architecture represented a form of nonverbal communication, as well as an earthly symbolic representation of the cosmos. This spiritual landscape was manifested at different scales across the region, from the specific construction of great houses, to the

building of long-distance roads, to the development of enclosed structures known as kivas.

Association with Kivas

As discussed above, most Ancestral Puebloan great houses are associated with kivas, which are either located inside the great house or are aligned with the great house at a nearby location. Some great houses also contain what are known as elevated kivas, or enclosed kivas, which are found at a second-story level or higher. In rare cases, a series of up to four enclosed kivas are vertically stacked on top of one another to create a tower kiva. In "An Anasazi Shrine in Chaco Canyon," A. C. Hayes and T. C. Windes (1975) have argued that tower kivas functioned as part of a signaling network. However, signaling alone does not seem a sufficient explanation, because a similar height could have been attained without the labor and time needed to construct the tower kivas. Rather, as Lynne Sebastian (2001) notes in "Chaco Canyon and the Anasazi Southwest," the kivas have a distinctive set of features and are remarkably uniform, indicating that more was involved in their construction than just for the purposes of signaling. This conclusion is further supported by ethnographic evidence, as contemporary Puebloan peoples use kivas for spiritual and religious purposes only and signaling has no association with these structures. In fact, the Casa Rinconada kiva dating to the middle Ancestral Puebloan period is aligned on a true north-south line, demonstrating that the cardinal directions are intricately tied to Ancestral Puebloan spirituality. Not only are kivas and specific buildings such as great houses aligned for spiritual purposes, but the spiritual landscape of the Ancestral Puebloan influenced the construction and alignment of almost all of their communities in general.

Looking at Pueblo Bonito as a whole, for example, it is evident that the entire village is astronomically aligned. Specifically, the wall that divides the pueblo into western and eastern

Originally built for safety and housing, kivas, such as the Casa Rinconada kiva in New Mexico, have been used for sacred purposes by the Ancestral Puebloans and their descendents. (David Muench/Corbis)

halves was constructed very nearly along the meridian, while the western half of the south wall is nearly a precise east-west line.

Landscaping Evidence

The spiritual importance of Ancestral Puebloan buildings and their specific alignment is not only reflected in above-ground architectural structures but also in the landscape surrounding the communities and specific earthworks within the communities. In fact, some of the landscape was manipulated to such an extent that scholars have identified mounds and other earthworks across the Ancestral Puebloan region. The most spectacular and convincing examples of Ancestral Puebloan earthwork architecture are found at Chaco Canyon and are the paired rectangular mounds in front of Pueblo Bonito. These mounds were defined and faced by massive masonry retaining walls and were provided with broad formal stairways that accessed a plastered horizontal surface at 3.5 meters above the original grade. J. R. Stein and Stephen Lekson (1992) have argued in "Anasazi Ritual Landscapes" that the space between the two mounds and between the mounds and the great house are formalized avenues or roads. These roads, in conjunction with the alignment of the mounds, formed a visual link between Pueblo Bonito and the "isolated" great kiva, Casa Rinconada. Specifically, the configuration of the space between the mounds extended the alignment of the eastern wall of the housing for Great Kiva A, which is part of the larger Pueblo Bonito structure, and projected it as a line of sight to Casa Rinconada, demonstrating that these two centers of importance were both spiritually and physically linked.

Stein and Lekson have suggested that landscape modifications such as roads, mounds, and other earthworks that are ubiquitous throughout the Ancestral Puebloan region represent a ritual landscape focused on great houses. Where they are present, earthworks such as mounds spiritually define the space in front of the great house, linking the great house to the rest of the Ancestral Puebloan world by means of the extensive road network. The relationship between mounds and roads has been fairly well documented, and most scholars agree on their significance in defining and linking the Ancestral Puebloan world. For example, at Haystack Great House, earthworks define the roads, with some lying parallel to the roads while others collectively bind the avenue, which encircles the great house. Likewise at Kin Hocho'i, the great house is completely encircled by a sunken avenue 15 meters in width. The perimeter of this avenue is defined by a berm that is a composite of several earthen masses. Stein and Lekson have convincingly argued that this berm functioned as a barrier between the great house and the great beyond, conceptually delineating a spiritual order built on the alignment of the mounds, roads, and buildings. The order that resulted out of this alignment created a spiritual landscape recognizable across the Ancestral Puebloan region and was a guiding influence across the Ancestral Puebloan period chronologically.

This conclusion is supported by the fact that numerous examples of roads have been identified that connect noncontemporary landscapes and architectural features. These compose some of the longest and most formalized segments of constructed roadway known for the Ancestral Puebloan.

For example, the Ancestral Puebloan pueblo of Ats'ee Nitsaa is separated from Kin Hocho'i pueblo by 2.5 miles of roads. Likewise, road number 1 at Navajo Springs ruins is intensively formalized for a distance of 1 mile and was arguably a constructed feature for 5 miles, where it appears to articulate with Rocky Point, a late Ancestral Puebloan ritual landscape on the south bank of the Rio Puerco near Sanders, Arizona. A similar roadway connects the middle Ancestral Puebloan great house of Casa Negra with Goodman Point, a middle to late Ancestral Puebloan center near Cortez, Colorado. Similarly, as noted by Arthur Rohn (1963) in "Prehistoric Soil and Water Conservation on Chapin Mesa, Southwestern Colorado," Cliff Palace and Spruce Tree House, both late Ancestral Puebloan centers in Mesa Verde, articulate with Far View House and Pipe Shrine House/Mummy Lake, middle Ancestral Puebloan centers by way of a pronounced "ditch" 4 miles in length. The Armijo Group in the Cebolleta Mesa district is focused on the Dittert Ruin, an apparent middle Ancestral Puebloan great house that is linked by roads to several structures and large kivas. In fact John Roney (2001), in "Prehistoric Roads and Regional Integration in the Chacoan System," has convincingly argued that the prehistoric roads were not constructed to facilitate transportation and communication, but rather were built for some other purpose. Since there were no transportation requirements for the large scale of the roads, it is reasonable to assume that the scale and elaboration of the road network was not motivated by simple economic function. Attempts to empirically demonstrate the roads' pragmatic purpose have met with difficulty, although it is fairly well agreed that the roads did fulfill a functional role in Ancestral Puebloan spirituality. As W. James Judge (1991) has noted in "Chaco: Current Views of Prehistory and the Regional System," the roads and their associated structures composed a visual communication network that was operative throughout the region. Furthermore, as suggested by Anna Sofaer and colleagues (1990) in "An Astronomical Regional Pattern among the Major Buildings of the Chaco Culture of New Mexico," certain roads may have been visible expressions of Ancestral Puebloan cosmology, supporting the argument that the alignment of Ancestral Puebloan communities was spiritually based.

Pictographic and Astronomical Alignment Evidence

Not only do roads and various architectural phenomena demonstrate the importance of spirituality in linking the Ancestral Puebloan landscape, but pictographic evidence also supports the conclusion that Ancestral Puebloan communities were aligned primarily for spiritual, and not signaling, purposes. Throughout the

Ancestral Puebloan region, pictographs of various designs link the buildings and communities to one another and to various astronomical phenomena. The most common astronomical phenomena marked by the Ancestral Puebloan were the solstices and equinoxes. One of the better known pictographs marking such an event is located a few hundred yards east of Wijiji Pueblo and is a symbol of the sun. Painted in white with what appear to be four feathers protruding from its disk, the sun symbol is located adjacent to an ancient roadway. Directly aligned with this symbol, 500 yards away across a shallow rincon, a sharp break appears in the mesa profile where a pillar of rock projects above the line of the horizon. This sandstone pillar is in perfect alignment with the painted sun symbol, and on the winter solstice the sun rises directly behind this sandstone pillar in line with the sun symbol.

Another example of specific Ancestral Puebloan astronomical alignment comes from the Holly House ruins, where two enormous boulders form a natural narrow corridor, 18 feet long, that opens to the eastern horizon. The western horizon is almost fully blocked by another large boulder. On the south side of this natural corridor, the Ancestral Puebloan pecked several shallow petroglyphs. Above and to the east are two spirals, and three feet to the right are three concentric circles with a dot in the center. The appearance of this symbol in this exact location clearly demonstrates sun watching was widespread and important for Ancestral Puebloan people. The spiritual importance of this site is further demonstrated by the fact that about 45 minutes after the sun rises above the local landscape on the solstice, a narrow sliver of light begins to enter the long, narrow corridor from a natural slit in the northeast and cut across the left-hand spiral. As it elongates and reaches across the second spiral, a second knifelike light beam begins as a tiny point of light far to the right of the sun symbol. This beam moves to the left toward the sun symbol and the first sword of light, eventually cutting across the sun symbol just below its center on the solstice. Likewise, at Fajada Butte a streak of light is caused by sunlight passing between two of the three upright sandstone slabs that lean back against the ledge on which a pictograph of a spiral is carved. On the summer solstice a small spot of light forms above the center of the spiral. It soon lengthens into a dagger shape and begins to move slowly downward through the center of the spiral. By paying close attention to the solar year, the Ancestral Puebloan could keep track of important times so that spiritual observances could take place.

Not only did the Ancestral Puebloan specifically record the solstices, but there is also clear evidence that they paid attention to other astronomical phenomena for various spiritual purposes. Perhaps one of the most sophisticated examples of Ancestral Puebloan spiritual alignment can be found at Hovenweep, located in southwestern Colorado. Within the Hovenweep area, as discussed by Ray Williamson in *Living the Sky*, Hovenweep Castle is a large tower that functioned as an observatory of the sun and contained the essential

elements of a complete yearly calendar marker. All four major solar events are marked by the building—the summer and winter solstices and the spring and fall equinoxes. Furthermore, the Ancestral Puebloan placed the external door in such a position that they could use it to keep a daily calendar along the northern interior wall.

In fact the variety of ways that the Ancestral Puebloan found to keep a calendar are truly astounding. They used ports, doorways, walls, and even the relative orientation of two buildings. For example, Unit-Type House, a boulder house located in the Hovenweep area, contains three solar ports opening, respectively, to the summer solstice, fall equinox, and winter solstices. By specifically aligning buildings throughout the region, the Ancestral Puebloan could keep track of the solar year and make sure that various rituals and observances took place synchonistically throughout the region.

Conclusion

Ever since Euro-Americans first came across the magnificent ruins of the Ancestral Puebloan peoples in the four corners region of the U.S. Southwest, hypotheses have been put forth attempting to explain the alignment of these ruins. Arguments that the Ancestral Puebloan communities were aligned for signaling purposes, although receiving initial support, are no longer tenable in face of the overwhelming empirical evidence. Architectural features such as the specific and planned construction of great houses; their connection to and association with great kivas, roads, and earthworks; as well as numerous examples of archaeoastronomic phenomena all overwhelmingly argue that the alignment of Ancestral Puebloan communities was for spiritual purposes.

References and Further Reading

Carlson, John, and W. James Judge. *Astronomy and Ceremony in the Prehistoric Southwest*, Vol. 2. Albuquerque, NM: Maxwell Museum of Anthropology, 1987.

DiPeso, Charles. *Casas Grandes: A Fallen Trading Center of the Gran Chichimeca*. Dragoon, AZ.: Amerind Foundation, 1974.

Doyel, David E., ed. *Anasazi Regional Organization and the Chaco System*. Albuquerque, NM: Maxwell Museum of Anthropology, 2001.

Doyel, David E., and Stephen Lekson. "Regional Organization in the American Southwest." In *Anasazi Regional Organization and the Chaco System*. Edited by D. E. Doyel. Albuquerque, NM: Maxwell Museum of Anthropology, 2001.

Ferdon, Edwin N. *A Trial Survey of Mexican-Southwestern Architectural Parallels*, Vol. 21. Santa Fe, NM: School of American Research, 1955.

Fowler, Andrew P., and John R. Stein. "The Anasazi Great House in Space, Time, and Paradigm." In *Anasazi Regional Organization and the Chaco System*. Edited by D. E. Doyel. Albuquerque, NM: Maxwell Museum of Anthropology, 2001,

Hayes, A. C., and T. C. Windes. "An Anasazi Shrine in Chaco Canyon." In *Papers in Honor of Florence Hawley Ellis* (Vol. 2, pp. 143–56). Edited by T. R. Frisbie. Santa Fe, NM: Archaeological Society of New Mexico, 1975.

Hewett, Edgar. *The Chaco Canyon and Its Monuments*. Albuquerque: University of New Mexico, 1936.

Judd, Neil M. "The Material Culture of Pueblo Bonito." In *Smithsonian Miscellaneous Collections 124*. Washington, DC: Smithsonian Institution Press, 1954.

Judge, W. James. "Chaco: Current Views of Prehistory and the Regional System." In *Chaco and Hohokam: Prehistoric Regional Systems in the American Southwest* (pp. 11–30). Edited by P. L. Crown and W. J. Judge. Santa Fe, NM: School of American Research, 1991.

Judge, W. James, and J. Schelberg. "Recent Research on Chaco Prehistory." In *Reports of the Chaco Center No. 8*. Albuquerque, NM: National Park Service, 1984.

Kelley, J. Charles, and Ellen Abbott Kelley. "An Alternate Hypothesis for the Explanation of Anasazi Culture History." In *Collected Papers in Honor of Florence Hawley Ellis* (Vol. 2, pp. 178–233). Edited by T. R. Frisbie. Albuquerque, NM: Archaeological Society of New Mexico, 1975.

Lekson, Stephen. "Standing Architecture at Chaco Canyon and the Interpretation of Local and Regional Organization." In *Recent Research on Chaco Prehistory* (Vol. 8, pp. 55–74). Edited by W. J. Judge. Albuquerque, NM: Reports of the Chaco Center, 1984.

Lister, Robert H. "Mesoamerican Influences at Chaco Canyon, New Mexico." In *Across the Chichimec Sea: Papers in Honor of J. Charles Kelley* (pp. 233–41). Edited by C. Riley and B. Hedrick. Carbondale, IL: Southern Illinois Press, 1978.

Rohn, Arthur H. "Prehistoric Soil and Water Conservation on Chapin Mesa, Southwestern Colorado." *American Antiquity* 28, 4 (1963): 441–55.

Roney, John R. "Prehistoric Roads and Regional Integration in the Chacoan System." In *Anasazi Regional Organization and the Chaco System* (Vol. 5, pp. 123–31). Edited by D. E. Doyel. Albuquerque, NM: Maxwell Museum of Anthropology, 2001.

Sebastian, Lynne. "Chaco Canyon and the Anasazi Southwest: Changing Views of Sociopolitical Organization." In *Anasazi Regional Organization and the*

Chaco System (Vol. 5, pp. 23–31). Edited by D. E. Doyel. Albuquerque, NM: Maxwell Museum of Anthropology, 2001.

Sofaer, Anna, Rolf M. Sinclair, and Joey B. Donahue. "An Astronomical Regional Pattern among the Major Buildings of the Chaco Culture of New Mexico." Third International Conference on Archaeoastronomy, University of St. Andrews, Scotland, 1990.

Stein, J. R., and Stephen Lekson. "Anasazi Ritual Landscapes." In *Anasazi Regional Organization and the Chaco System* (pp. 87–100). Edited by D. E. Doyel. Maxwell Museum of Anthropology Anthropological Papers 5. Albuquerque, NM: University of New Mexico, 1992.

Van Dyke, Ruth M. "Memory, Meaning, and Masonry: The Late Bonito Chacoan Landscape." *American Antiquity* 69, 3 (2004): 413–31.

Williamson, Ray A. *Living the Sky: The Cosmos of the American Indian*. Norman: University of Oklahoma Press, 1984.

Windes, Thomas C. "Population Estimates." In *Summary of Tests and Excavations at the Pueblo Alto Community. Investigations at the Pueblo Alto Complex, Chaco Canyon, New Mexico, 1975–1979* (Vol. 1, pp. 383–406). Edited by T. C. Windes. Santa Fe, NM: National Park Service, 1987.

3

The Mayan kingdoms died out from disease. (Earlier views held Mayan decline to be cyclical.)

PRO   Justin Corfield
CON   Chris Howell

PRO

The Mayan civilization was one of the greatest in the medieval world. At its height it dominated southern Mexico, the Yucatan Peninsula, and also large regions of Central America, including parts of modern-day Belize, El Salvador, Guatemala, and Honduras. There is archaeological evidence for the Mayan civilization starting as early as 2000 BCE, and the Mayans are noted for having the only known fully developed written language in pre-Columbian America. Their civilization reached a peak during what is now called the classic period of the Mayas, from about 250 to 900 CE. However, parts of the civilization went into decline in the 8th and 9th centuries, and by the start of the 11th century, the civilization had decayed, leaving behind incredible ruined cities and evidence of urbanization on a massive scale. The city of Tula seemed to have already fallen into major decline, although the exact date of its ending remains obscure. The city of Chichén Itzá continued to flourish until about 1250, when it was finally sacked by an army from Mayapán.

Some archaeological work at El Mirador in modern-day Guatemala has shown that the decline in that part of the Mayan world started as early as about 300 CE, but for the most part the decline was not until much later. There is no written documentation explaining the end of the Mayan civilization. Instead, the decline is dated from the lack of written sources. This has led to many suggestions that the Mayan civilization ultimately collapsed as a result of disease, which possibly destroyed the cities, forcing people back into the countryside. This may explain the abandonment of the cities, but that alone cannot explain how their civilization collapsed steadily over such a relatively long period of time. In recent years there have been a number of studies of the climate at the time, and many scholars now suspect that environmental problems, brought on by a long drought, were responsible for the decline of the Mayan civilization. This in turn seems to have led to a series of disagreements between people in various cities, leading to the eventual collapse.

Petroglyphs

One of the oldest methods of creating lasting images, petroglyphs have been made all over the world, since prehistoric times. Instead of drawing or painting on a stone surface, or carving a larger stone into a sculpture, a petroglyph is made by chipping away at a stone surface in order to engrave the image. Various methods can be used, depending on the type of rock and the desired image, and modern people are often surprised to see how detailed and skillful petroglyphs can be. They are often discussed because of the similarity of petroglyph images at sites around the world, which is sometimes used to support various diffusionist theories of history or Jungian theories of psychology.

Rise of the Mayan Civilization

To analyze the fall of the Mayan civilization, it is first important to recognize why the civilization became so powerful. Although archaeologists have found remains in the Mayan area going back to the 10th millennium BCE, these are not clearly identifiable as being "Maya" until sometime from about 1800 BCE, in the Soconusco region of Mexico, near the Pacific Coast. The pots and the clay figurines from that period clearly appear to be precursors of later Mayan items, and by 1000 BCE, these "preclassic" Mayas were using burial mounds that have some similarities with the subsequent stepped pyramids for which they became famous. These early sites are located in the southern Mayan lowlands, with locations in the Guatemalan highlands being settled from about 800 BCE. Gradually, Mayan communities started to be established in the northern Mayan lowlands.

During the classic period of Mayan civilization (from about 250 to 900 CE), there was a rapid expansion of Mayan settlements, with large settlements being transformed into cities, new towns being built in a large number of locations, and other settlements developing into towns. There are also considerable surviving monumental inscriptions and also significant intellectual and artistic achievements evident in surviving temples and other buildings. The great city of Tikal and also the cities of Calakmul, Copán, and Palenque all date from this period. The largest monuments in these cities are the large stepped pyramids and, also nearby, what appear to be clearly the palaces of the Mayan rulers. The rulers themselves were more like mayors or tribal chiefs, rather than the royalty of medieval Europe. They ruled with the people and clearly had considerable interactions with them.

Start of the Decline

The surviving inscriptions at Tikal and elsewhere describe the great achievements of the rulers and the Maya in general, as well as showing clearly that there was

widespread trade in the region. Certainly, archaeologists have found gold at Mayan sites that originated in Panama and also goods that have clearly been brought in from long distances. Some of the achievements of the Maya exceeded those of western Europe during the same period, and some of the Mayan settlements were larger and more complex than those in Europe. However, during the eighth and ninth centuries the Mayan civilization declined quickly. Gradually, by about 800 CE, there are fewer and fewer inscriptions. There also were not many large-scale buildings being constructed, and many of the cities went into decline and were subsequently abandoned. Some writers have suggested that the Mayan civilization did not collapse so quickly. It was, they have argued, more likely that there was no new building work, and the same people and their descendants continued to live in a city but without the money to build more, or to repair it, and that the abandonment of the cities came later. Either way, there was a gradual decline in the whole region—the only matter in dispute is the exact speed of the decline.

Comparable Civilization Collapses

The Maya were not the first civilization in Mesoamerica to collapse—the Izapa operated in the Mexican state of Chiapas and operated from 600 BCE until 100 CE. They operated before the Maya, with many surviving temples, inscriptions, and also stelae. The reason for their sudden decline is also not known, and it is possible that the Mayan civilization could have collapsed for similar reasons.

Some civilizations around the world have been brought down by a single cat-astrophic or cataclysmic event, and there have been suggestions that this might have been the case with the Maya. There are active volcanoes in the region, and there certainly were some in the 9th and 10th centuries. However, there are three problems with the "single catastrophe" theory to explain the collapse of the Maya. The first is that there is no clear evidence of any major volcanic eruption during the period of the decline; the second is that a volcano would probably only strike one center, not all of them (and the Mayan civilization covered a large area); and third, the decline of the Mayan civilization was gradual, taking place over more than a century. By comparison with the Romans, devastating as the eruptions of Vesuvius for Pompeii and nearby places have been, the city was relatively quickly rebuilt after the eruption in 62 CE, and even after 79, settle-ments were built nearby even if the original city, then covered in lava, was no longer inhabited.

Rise of Disease as Reason for Mayan Decline

The lack of any convincing evidence for the "single catastrophe" theory has led some people to contend that a disease could have been responsible for their decline. There are many diseases prevalent in parts of Mexico and Central America

today, and visitors to the country are always advised to take precautions against malaria in some parts of the country, as well as dengue fever, schistosomiasis (bilharzia), typhoid, and tuberculosis. Certainly there were many scholars who surmised that a credible way of explaining the end of the Mayan civilization must have been because of disease, including a possible poisoning of the water supply. It seemed as though it was the only possible explanation for the relatively sudden collapse of the cities, with the population continuing in the countryside. These ideas were certainly influenced by the study of the history of disease in the 19th and early 20th centuries, with many books written, and even more studies and surveys conducted into the effect of the Black Death, which hit Europe in 1340s, as well as the other plagues that routinely devastated parts of Europe.

Epidemiologists have studied how disease has been able to devastate a population. Because of the tropical climate, which would help disease to spread, and the close confines of many Mayan peoples in cities, the explanation that the Mayan civilization died out from disease became accepted by many scholars, and their theories were then repeated in more general works on the Maya. It was also suggested that some other civilizations, including the empire of Angkor in modern-day Cambodia, might also have been sapped by disease, although its final collapse was after its capital was sacked by the T'ais from modern-day Thailand.

Although there is no evidence that any diseases have been capable of killing the entire population of a particular city, it would not have been necessary for this to happen to cripple the Mayan civilization. Diseases may affect particular people in a city more rapidly, with some modern diseases often affecting different segments of the population much faster than others. Had a virulent disease struck the Mayan civilization and killed many of the intelligentsia and the middle class as well as those living in close confines in the city slums, it would easily have been possible to cause a quick end to major building projects (through shortage of labor if nothing else) and a rapid decline in the civilization. In *A Study of History* (1934), the British historian Arnold J. Toynbee suggested that the "creative minority" of the population was the central part in any civilization, and the deterioration of their role could have a more far-reaching effect than the death of a percentage across the entire population. As a result, if there were a disease that was rife within the elite in a city—such as wealthy Romans suffering from poisoning owing to their reliance on water from lead pipes—this certainly could account for a decline. Yet, when this happened in other places, it was not long before a new elite emerged.

Another reason why disease may not be the sole explanation for the collapse of the Maya was that during events resulting in the death of large numbers of people, such as in the Black Death in Europe in the 1340s (which is believed to have killed between 30 and 60 percent of the entire population of the countries ravaged), there was far less demand for food, less need to cultivate marginal land, and more resources per head for the people who survived. This would

allow, in time, the population to rebuild itself. This certainly happened in late medieval Europe. But this "rebuilding" did not take place with the Maya. Although the plague did return in Europe at intervals—often once every generation, or so—and although tens of thousands of people were killed in these pandemics, there was never any risk of the destruction of any entire nation. Even the Spanish influenza epidemic that followed World War I (1914–18), breaking out around the world from 1919 onward, which caused a similar number of deaths to those who had died in the war, never came close to wiping out entire populations, even on isolated islands in the Pacific, where it struck with devastating effects on American Samoa and New Caledonia.

Environmental Issues as Cause of Mayan Demise

Instead of disease as a major factor, it now appears that there were major environmental problems whereby the Maya had clearly overexploited the land. It is not known what the exact population of the Mayan kingdoms was, or even that of individual cities, although rough (and varying) estimates have been made. Like most cities around the world, they were probably on the site of small townships, located in the middle of fertile land, and the townships grew into cities, the sprawl covering much of the fertile farming land, thereby eroding the farming base of the region.

Environmentally, but on a very different scale, the demise of the classic Mayan civilization can be compared to the events on Easter Island, where the overpopulation has stripped the environment, and where a drought could quickly cause major problems that some people chose to try to resolve in war. This does seem to explain the early demise of the Mayan settlements at El Mirador. Certainly, the Mayan god Chaac, the god of thunder and rain, was one of the most important gods in the Mayan world. There were a number of rituals associated with him, and he was the patron of agriculture, represented with a human body covered in reptilian scales, and a nonhuman head, which had a long and pendulous nose and fangs. Although he was an important Mayan deity, not too much need be read into this, as many other cultures around the world had gods of rain—the Aztecs having Tlaloc—because the agricultural cycle was important to all civilizations.

Even without a drought, there were enough environmental problems in the Mayan Empire. Around the Yucatan Peninsula and the Peten Basin, there has long been believed to have existed relatively thin tropical topsoil. The removal of the forest cover generally results in quick degradation of the land, and it seems likely that many of the trees would have been felled for construction work in the building not only of the surviving stone temples but also the many wooden houses that must have been residences for ordinary people in Tikal and other cities. There was little ground water for the Maya, and there are also not many sources of fresh water in the Yucatan Peninsula, which has few lakes. There are no river systems

Tikal, in present-day Guatemala, was a hub of Mayan culture during the Classic Period (CA. 200 to 900 CE). It boasted a population of about 100,000 people. The excavated ruins cover six square miles, but most of the vast city remains uncovered. (iStockPhoto)

in the Peten Basin. A regular monsoon rain is required to keep the tropical vegetation alive, and when this does not happen through drought, it would have been possible for a collapse in the agricultural base of the country to happen quickly.

Spanish colonial officials were able to document these cycles of drought, and it is believed that these patterns might have happened earlier. Indeed, some climate change skeptics have argued that global warming is a regular feature of the Earth, and that the rise of the climate that led to the drought that ended the Mayan civilization might have just been one of these cyclical changes.

The Maya used a very intensive cultivation system, which used ridged terraces, and also in some areas, they clearly used slash-and-burn "techniques" for agriculture. The latter had the effect of making land good for use soon after it was deforested, but the soil could rarely sustain crops for long periods of time. The terracing produced the result of keeping the topsoil, but the intensive nature of the planting cycle meant that natural fertilizers had to be used to keep up the high levels of production. If, in periods of "plenty," there were large increases in population, then any agricultural problems would lead to starvation. However, as with diseases, it would not necessarily lead to the total collapse of the civilization, unless there were some other factors.

The two Mayan scholars Thomas Gann and J. Eric Thompson first suggested the idea of a drought leading to the collapse of the Mayan civilization in

1931. Irishman Thomas W. F. Gann (1867–1938), a medical doctor by training, became interested in the Maya during his posting as a doctor to British Honduras (modern-day Belize). Although he was at first fascinated by the concept of a civilization dying out through disease, he gradually came to focus on studying the possibility of a drought and how that could affect the region. J. Eric Thompson (1898–1975), an English archaeologist who also became involved in studying the Maya through his time in British Honduras, was involved in deciphering the Mayan script.

Both Gann and Thompson felt that a drought was possible, and the whole concept was explained in much more detail by Richardson Gill (2000) in his book *The Great Maya Droughts*. Gill felt that it was a long series of droughts that started the problems leading to the end of the Mayan civilization. He argued that all the other possible causes, such as internal warfare, peasant revolts, a fall-off in trade, and foreign invasion, could be explained by a shortage of food resulting from drought. Drought could lead to an abandoning of particular cities and a gradual but inevitable decline in the Mayan civilization.

As to examining whether or not there was a drought at the time, sediment cores from Yucatan have provided evidence for a severe drought from 800 to 1000 CE, which was the most severe in the last 7,000 years and which coincides with the fall of the Mayan civilization. A study of temperatures around the world has shown that cold weather in northern Europe usually coincides with drought in Mesoamerica, and a study of climatic models, tree-ring data, and surviving written accounts shows that there was a cold spell in northern Europe during the period of the Mayan collapse.

The Maya clearly saw that there would be a problem with a drop in the water supply. They lived in a seasonal desert, and rainwater was their main source of water both for drinking and for irrigation. From winter until spring, there was little rain, so that during the rest of the year, water had to be preserved. Sometimes natural depressions were lowered to help with the construction of reservoirs, with many houses building their own small reservoir, or *chultun*, in the ground, sometimes actually underneath the house or nearby. Certainly, some Mayan settlements, such as those around the Petexbatun Lake and also others along rivers, did have regular supplies of water. But in cities such as Tikal, the population had to survive entirely on rainwater. There were large water tanks built to collect and store the rainwater, and at times when the water ran out, there were clearly major problems. In fact, much of the building work at Tikal ended at the end of the eighth century, and this would mean that Tikal possibly was the first of the major Mayan settlements to be abandoned.

Some scholars have argued that the drought theory may easily explain the abandoning of Tikal, but it would not explain how northern cities such as Chichén Itzá, Uxmal, and Coba managed to survive and thrive. Gill, in his research, showed that groundwater around Tikal was too far down for the Maya to reach,

whereas groundwater in the north is much closer to the surface, and bore holes did not need to be as deep. Indeed, the U.S. archaeologist Sylvanus Griswold Morley (1946) showed that most of the cities that survived longer had a large number of these bore holes and had greater access to water. The coastal Maya were also able to flourish on heavier reliance on seafood to supplement their diet.

Argument for Political Origin of Mayan Demise

Archaeologists have discovered that the peak of the building projects for the Maya was between 730 and 790 CE. In any society of the period, the majority of the building work had to be undertaken by the peasants. This has led to some left-wing scholars arguing that the Mayan civilization could have been brought down by a peasant revolt. There is no written evidence for a peasant revolt, but the absence of some details on inscriptions suggests that there could have been a number of uprisings. In many other comparable civilizations, where there are much better written records, there were revolts, and the Maya would not have been an exception. There is some archaeological evidence for the revolt through the smashed remains of some thrones and the burning of some of the temples, events that remain characteristic of a peasant revolt or revolution. If these revolts had taken place, the elite would gradually have lost control of the population, and with revolts breaking out around the Mayan Empire, it was possible that the empire might have decayed over a couple of centuries through societal collapse.

This theory is not incompatible with the idea of a drought. If a drought had taken place, and food shortages ensued, the civilization would have been at an increased risk of a peasant revolt. The elite kept their position essentially by providing the necessities of life to the poorer people. If they were unable to provide food, or were not able to supply water, people would quickly lose faith in the governing class, and it is inevitable that there would be some attempt at a revolt. It might also have been possible for some Maya in some areas to join up with people in nearby cities or towns to usurp the rule of their elite at periods of crisis. Again, this is not incompatible with the theory of a prolonged drought.

Some theories of history hold that great social turmoil, revolt, or revolution usually stem from foreign wars. Certainly the French Revolution of 1789, the Russian Revolutions in 1905 and in 1917, and the Chinese Revolutions of 1911 and 1949 owe some of their origins to other wars. Fighting is known to have taken place between the people from Chichén Itzá, who attacked and captured their southern rival Seibal in 830 CE, which then allowed them to continue their attack on Becán. There is also evidence that the Toltecs did attack the Maya in Yucatan. But there is no evidence of any major invasion of the Mayan lands. The destruction of some places through fighting could, again, go some way to back up the theory of drought. In times of crisis, the inhabitants of one city might invade another to seize its resources, especially water. If the Mayan system was crumbling, the fact that it did

not face a widespread invasion by the Toltecs could be explained by the Maya having declining resources and abandoning areas as not viable for their population, and therefore not worth the Toltecs investing their time and trouble to invade.

There is evidence that there was considerable fighting within the Mayan kingdoms, with the kingdom of Mutal, based at Tikal, and that of Kaan, based at Calakmul, fighting in the sixth and seventh centuries, respectively. Fighting tended to take place between rival kingdoms, and an increase in fighting could exacerbate the environmental degradation. However, the fighting in the Mayan world just before its decline, for which there is evidence, shows that it was either dynastic disputes between rivals or between different city–states. Some historians claim this alone could not result in the collapse of the entire civilization, pointing out that even the devastation of parts of Europe during the Hundred Years' War in France or the Thirty Years' War in central Europe did not result in the permanent abandonment of the cities attacked. Nevertheless, there are many instances in history when this has been the case, such as the sacking of the Cambodian city of Angkor in 1432 and the destruction of the T'ai city of Ayuthia in modern-day Thailand, both of which led to the end of their respective civilizations. Part of this has been explained by the British historian Arnold Toynbee, who observed, after studying the fighting in Ceylon, that a war in a tropical region can be more devastating than that in a temperate zone. This is because the destruction of the methods of keeping water such as the irrigation system or the water supply system is far more devastating than in temperate climates, where rainfall could replenish the water supply.

The fighting and the peasant revolts certainly disrupted the trade routes that had made the Maya wealthy, even though they did not use beasts of burden and had not invented the wheel. However, most of the long-distance trade during the period of the Maya was in luxury items such as gold, feathers, cacao, and other items that could be carried relatively easily over long distances. Food was traditionally grown outside cities, in close proximity to the centers of population, and thus any disruption in the trade routes, however sustained, could not lead, in itself, to the collapse of the Mayan civilization. Some trade routes between Teotihuacán, Monte Albán, and Mayaland were disrupted in the seventh century CE, with the quality of ceramics and other goods seeming to decline significantly in quality. Certainly, archaeological work on Mayan sites in the northern Yucatan Peninsula have shown that some of the settlements there seem to have taken over the trade routes of their former rivals farther south, but this would be no more than a reflection of the decline, not a cause of it.

Conclusion

All the scenarios for understanding the end of the Maya basically surround a series of tumultuous events that affected everyone in the Mayan kingdoms. It had to be gradual and sustained—because the Mayan civilization collapsed slowly

but inexorably. This returns us to the hypothesis that it was a great drought, brought on from climatic change, that left many of the Mayan cities, such as Tikal, without supplies of water. This, in turn, led to an end of building projects as the population started to become unsettled. Following this were social upheavals, and these led to peasant revolts and attempted revolutions. The cessation of construction of new temples and prestige projects seems to be clearly, according to the archaeologists, before the time of the temple burnings, and this scenario therefore fits with the available evidence.

The civil unrest led to the inhabitants of some cities attacking those of others, a splintering of central rule, and disunity, because those cities with large water supplies could maintain themselves and those without had serious problems. However, the result was that the inhabitants of those cities without water would attack those with, and this, in turn, could lead to a series of internecine wars in which the environment—the water supplies, the farms, and the method of producing and selling food—was so severely disrupted over so much of the country that the civilization collapsed on itself. The invasions by other powers seem to have taken place when either the Maya were very weak or after their civilization had already collapsed.

References and Further Reading

Coe, Michael D. *The Maya*. London: Thames and Hudson, 1999.

Demarest, Arthur Andrew. *Ancient Maya: Rise and Fall of a Rainforest Civilization*. Cambridge: Cambridge University Press, 2004.

Gann, Thomas and J. Eric Thompson. *The History of the Maya from the Earliest Times to the Present Day*. New York: C. Scribner's Son, 1931.

Gill, Richardson Benedict. *The Great Maya Droughts: Water, Life and Death*. Albuquerque, NM: University of New Mexico Press, 2000.

Henderson, John S. *The World of the Ancient Maya*. Ithaca, NY: Cornell University Press, 1997.

Hernandez, A. Arellano. *The Mayas of the Classic Period*. Woodbridge, England: Jaca Books, 1999.

Morley, Sylvanus Griswold. *The Ancient Maya*. Stanford, CA: Stanford University Press, 1946.

Webster, David L. *The Fall of the Ancient Maya: Solving the Mystery of the Maya Collapse*. London: Thames and Hudson, 2002.

CON

Although several million people of Mayan descent are still with us today, a certain fascination with the classical Mayan world, including its tropical cities, highly accurate astronomical observations, hieroglyphic scripts not fully deciphered, and

the mysterious collapse of the classic Maya over a millennium ago, continues to captivate the imaginations of scholar and layperson alike. What happened to the millions of classical Maya occupying dozens of stone cities in the heart of a Central American jungle? Well, of course today, through archaeology, linguistic advances in the ancient scripts, and through the modern Maya themselves, we know that their civilization did undergo a collapse of sorts in its core, the southern or Peten lowlands by 900 CE, although the Maya did not disappear but simply abandoned cities and temple complexes that could no longer be sustained, often migrating to peripheral regions. After 600 years of successful civilization building, what happened?

As a research community we can now answer many of the who, what, when, and where questions of the classical Mayan collapse, but to date, the exact nature of the collapse, or as some would prefer, the transformation from a classical to postclassical world, including "how" and "why" the collapse took place, cannot be answered. Over 80 theories have been put forth as to what caused the collapse in between 800 and 900 CE. Perhaps the best summary to date was created by Heather McKillop in 2004. McKillop breaks down the explanations of classic Maya, Peten lowland collapse theories into three main categories: ecological overload, endemic warfare, and climatic catastrophe. David Webster (2004) has argued it was ecological collapse, while Arthur Demarest et al. (2005) have argued for the endemic warfare theory and against the other two theories. Interestingly it was Demarest's team that researched the collapse of an entire Peten region known as Petexbatun, centered on Mayan cities like Dos Pilas and Aguateca, that led Demarest (1998) and others to the endemic warfare conclusion. Not only did these researchers find evidence of a violent, warfare-driven collapse of the region, but they did not find evidence for ecological collapse or climatic change. This research project is probably the most convincing to date arguing for the warfare theory. David Webster (2004), though a proponent of human driven ecological change as part of the collapse, took the warfare argument a step further to include the entire southern lowlands for late classic Mayan times. He argued for a combination of internecine warfare, royal rivalry leading to further warfare, and an exploding population leading to ecological imbalance and gradual decline. Together, these arguments differ mainly in scope and in rate of change or decline rather than in any fundamental way that excludes warfare from the equation.

The climatic change theory has been repeatedly postulated and then fallen from favor. In 1931 Thomas Gann and J. Eric Thompson, long the preeminent Mayan researchers, suggested climate change and social revolution. In modern times, research by Richardson Gill (2000) has renewed interest in climate change as the major cause. The problem with this theory is that most of the Gill data are from lake cores in the northern lowlands of the Yucatan and not from the core collapse area, the southern lowlands, or Peten. McKillop (2004) has noted sea level rise in a coastline near the Peten. However, without comprehensive

research in the core collapse area of the Peten itself, along the lines of what took place with Demarest's team in the Petexbatun, it's difficult to explain why a major drought in the Yucatan would have led to increased Mayan growth there from 800–1000 CE while the region to the south of Gill's database collapsed! Further, why would a climate change only strike the Peten and not the surrounding regions? The Maya in Copan in highland Honduras continued on a slow decline, while the great city–state to the north in central Mexico, Teotihuacan, had already been abandoned by 650 CE. Also the so-called great drought of 800–1000 CE, supposedly the greatest in the past 7,000 years, does not explain the decline of Mayan cities along waterways such as the Usumacinta, Pasion, Negro, or Belize rivers. The classic Mayan cities of Yaxchilan and Piedras Negras lie along the banks of the Usumacinta at the edge of the Petexbatun project. Classic Mayan sites were abandoned along some rivers while others grew.

Jared Diamond (2003, 2005) has also suggested environmental collapse, but such an argument is subject to the same critique leveled above. Diamond (2003) does note that classic Mayan warfare peaked just before the last recorded Mayan calendar long count date of 909 CE and counts warfare as one of his five most important factors in assessing the collapse. Diamond (2005), like Webster and Demarest, suggests the collapse is a complex issue and not easily explained by a single climatic change theory. All researchers acknowledge that warfare was a constant part of Mayan civilization, whereas environmental change, whether human or naturally derived, was more periodic. Researchers also find a fundamental issue that is hard to explain. Why did the Maya not only abandon but also not return to the Peten even after the forests returned and the drought ended? Was there something more than environmental issues that kept them away? Was this a buffer zone both in time and space from a land of trouble?

The classic Mayan collapse is complex. The Maya utilized a variety of food production and water control measures exactly because they were in a classic wet-dry rainforest biome. The very nature of the civilization developed around this ability to adapt to seasonal change in the rainforest. Additional complexities include the uneven collapse, where core areas of classic Mayan civilization in the Peten are abandoned while peripheral areas are slower to decline or even increase in population. Taken together, this complexity of the case makes it unlikely that a single catastrophic theory will explain the collapse. Demarest has repeatedly argued just such a case and additionally has carried out the most pointed research looking for evidence directly within the collapse region of all three main explanations of ecological change, endemic warfare, and climatic change.

Certainly much of the debate centers on what one means by "cause" and by "collapse." Demarest and Don and Prudence Rice (2005) have discussed the complexities of labeling historic change across many variant regions that made up the Mayan world. It will suffice here to define a working definition of "cause" as meaning the leading reason or reasons and "collapse" as referring to a major set

of changes concerning how a civilization survives and thrives. It will be argued that although population excess, resource depletion, and climatic change were factors or reasons in the classic Mayan collapse, warfare, in a broad sense, was the frontrunner in bringing about the end and transformation of the classical Mayan world for both commoner and elite over 1,000 years ago.

Background to the Controversy

The Maya are a set of peoples who shared 30 related languages across five major time periods and three geographic regions. Temporally the Maya start their calendar on August 11, 3114 BCE, but archaeologists cannot find the first signs of distinctive Mayan culture until the last millennium BCE. Thus archaeologists date the preclassic Maya from 1000 BCE to about 200 CE. The period is poorly understood currently due to lack of excavations and Mayan scripts that date to the time period. However, sites like El Mirador indicate just how extensive the Maya were at this time. The classic Maya existed from roughly 200–1000 CE and are better understood due to more archaeological and historical research. This period is characterized by stone cities, temple complexes, complex calendars, murals, and hieroglyphic scripts that vary by region but as a rule are much better understood now than they were 20 years ago. Major cities like Tikal, Caracol, Palenque, and Copan vied for power as the Mayan populations reached their zenith. The postclassic Maya existed from 1000–1500 CE and are often associated with a move of population and power north to the Yucatan Peninsula at sites like Chichén Itzá and Coba. Eventually the arrival of the Aztecs, and on their heels, the Spanish, by 1500 CE truncated this reemergence of the Maya. This led to the historic era of Mayan civilization, one in which the many Mayan communities shrank from the diseases and conquests of the Columbian exchange but have rebounded to populations approaching those of the classic period.

Geographically, the Maya occupied three regions beginning with the highlands in the south, where most Maya live today, namely in Guatemala and Honduras. Mayan population grew substantially in the jungle lowlands known as the Peten, north of the highlands by the start of classic times. Then in postclassic times, after the "collapse," more and more Maya migrated into the Yucatan limestone lowlands of the north. Thus we can think of three major Mayan regions: the southern highlands, the Peten lowlands in the middle (often referred to as the southern lowlands), and the Yucatan lowlands in the north. They relied heavily on corn for up to 70 percent of their dietary needs in an area roughly the size and population of the modern U.S. state of Colorado. Mayan population density averaged about 500 people per square mile according to Jarod Diamond (2005). Such numbers indicate how likely conflict due to population and resource access must have been, and Diamond notes that the Maya were constantly at war within themselves.

The Evidence

So what is the evidence that warfare was a primary factor in the classic Mayan collapse and not other factors, such as the oft-cited droughts or overharvesting of resources such as limestone or timber by a too large population? A combination of epigraphic (written) and archaeological evidence, much of it surfacing within the past 10 years, now strongly points toward large-scale warfare, both polity to polity and internal civil wars, as a major reason for the Mayan collapse.

Epigraphy (Written and Art)

Previous arguments against war's role in the Mayan collapse have centered on the limited ability of the Maya to wage war beyond a ritual clan level due to technological limitations, a peaceful nature, and small polities. Lawrence Keeley (1997) has argued that Stone Age warfare can be every bit as lethal as modern warfare in his groundbreaking work *War before Civilization*. Of the 15,000 plus known inscriptions, most fall within their Baktuns 8 and 9 or the classic period. Linda Schele and David Friedel (1990) have noted that Maya writing and art as well artifacts indicate a well-developed set of Stone Age tools for war, including spear throwers, obsidian-edged weapons, industries for producing weapons, trained warriors that utilized them, and perhaps most important, reasons for use. Webster (2004) believes those reasons included fulfillment of rituals, status rivalry, and most important, competition for resources. Further, the idea that the Mayan polities never attained sizable state parameters and thus could not carry out large-scale war was first attacked by Joyce Marcus in 1976 and was further elaborated upon by Simon Martin and Nikolai Grube in 2000. Martin and Grube believe the epigraphic evidence is strong enough to specifically define territorial states and boundaries of which alliances between Tikal and Naranjo and Calakmul and Caracol were perhaps key in understanding why the classic period rivalry between giant cities Tikal and Calakmul may hold clues to the Mayan collapse.

The power of elite status rivalry by royal families within the polities cannot be underestimated. As Ronald Wright (1992) has noted, the postclassic Mayan text known as *The Annals of the Cakchiquels* holds clues to the Mayan collapse. The Cakchiquels were a group of royalty who broke off from the largest Mayan nation of today, the Quiche, and became their rivals in 1475 CE. This took place not because of too few resources but because an opportunity and resources to form a rival polity were available. War was the result of this splintering. Such a sequence may be relevant for the incessant splintering and alliances noted in the late classic art and epigraphy of the lowlands as well.

In any case, Diane and Arlen Chase (1998) have compiled four types of warfare listed in classic Mayan inscriptions, including capture events, destruction events, decapitation events, and "star wars" (battles timed to correspond with astronomical events) usually involving the planet Venus. Linda Schele and

A wall painting from Bonampak depicting Maya priests and nobles judging prisoners of war. (Peabody Museum, Harvard University, Cambridge, MA/The Bridgeman Art Library)

David Friedel, David Stuart, and David Webster have all noted that the four types of warfare, and especially the star wars events, increase dramatically in late classic texts and art to become the dominant theme by 800 CE. The Tikal-Calakmul rivalry and associated alliances are connected with most of those events. Interestingly none of the over 15,000 inscriptions has been interpreted as describing times of drought or ecological disaster, especially those from the late classic. Warfare is the clear and dominant theme.

If the Chases and Martin and Grube have correctly interpreted a badly eroded monument, we even know the names of the heads of the royal families who perhaps engaged in the star wars episode that placed warfare in an exalted context as the primary means by which power in the region could be manipulated. It was in April 562 CE when Caracol ruler Yajaw Te' K'inich II, on behalf of Calakmul, defeated Tikal leader Wak Chan K'awiil. So while the exact names of clans and kin groups may not be known, the main protagonist polities and leaders certainly are.

Archaeology

It is important to remember that archaeology data for war are often hard to find or ambiguous due to their ephemeral nature. The research of Connell and Silverstein (2006) on the borders of territories between the powerful states of Tikal and Calakmul has yielded a surprising number of both hasty fortifications and

---

Core and Veneer Architecture

The architecture associated with the Puuc region of the Yucatan, core and veneer architecture uses a concrete core instead of the larger, bulkier stones that had been used earlier. This stable core is then decorated with a veneer of intricately cut stones placed before the concrete has set. The decorative veneer can vary from simple repeated patterns of shapes to masks to painstaking mosaics to elaborate sculptures. Examples of the style among the Mayan ruins can be found at Chichén Itzá and Oxkintok.

---

more permanent defensive citadels dating to the late or terminal classic including Tikal, Calakmul, and the associated polities Dos Pilas and Chunchucmil. It may be that archaeologists have been looking too much at the big temple complexes and not enough in the high conflict borders. The Sacbeobs, or raised causeways that emanate from places like Calakmul, may be key in determining key alliances and also crucial border areas, but much more work needs to be done. Archaeological research designed to test for warfare at secondary centers such as Caracol and Dos Pilas has been conducted by the Chases and by Arthur Demarest, respectively. Additionally, Demarest has found evidence of warfare in the late classic across the Petexbatun region.

Perhaps most telling is the congruence that takes place between star wars events, archaeological evidence of war, and the epigraphic data. The archaeological evidence for war is strongest both at major and minor polities associated with Tikal and Calakmul just when the star wars events are mentioned, most often in the Mayan glyphs and in Mayan art forms during the late postclassic. The grandest period of monumental architecture projects was in the ninth century CE as well. Classic Mayan civilization was at its height just before its collapse. This does not make it a certainty, but it is interesting to note that the Maya were not turning to new agricultural practices, new lands, or even to the supernatural powers as times got tougher, but instead they increased the level and intensity of warfare. The site of Seibal does have some indications of foreign influence from the north, but Demarest as well as Webster believe most of the warfare was endemic or internal between Mayan polities, and perhaps in some cases it broke down into civil unrest.

Conclusion

Perhaps the most important point to make about the role of warfare in the classic Mayan collapse concerns the interrelationship between population, environment, civilization, and war. No one doubts that Mayan populations were high, pushing the limits of their highland and lowland resource bases, and were exacerbated by

droughts, some severe. Yet the Maya had weathered previous droughts, successfully increased food production and water control in previous times of need, and some even migrated north into Yucatan lands during the postclassic, which was only lightly touched on by the conflicts of the Peten and the south. So if a region-wide drought was the key to the collapse, why did some areas like Copan and the Yucatan continue to thrive?

However, if war was the key theme in the late classic, and the epigraphic and archaeological evidence suggests it was by 800 CE, then the data showing that the Peten regions succumbed while others, especially peripheral regions, blossomed make more sense. An increasingly overpopulated core continued to grow by using alliances with smaller polities that provided access to resources not available at the core. These alliances were cemented by marriages, kin relations, common resource territories, and strategic locations to acquire needed materials such as obsidian, limestone, timber, exotics, trade route access, and slaves. The big polities were the military elite, and they provided coercive force support to the smaller polities in their local and border rivalries as the Peten lowlands became circumscribed. Seasonal warfare known to us as capture, destruction, and decapitation events more or less kept the balance of power in the early classic. But, in the late classic the balance of power changed, especially after the 562 CE fall of Tikal to Calakmul, Tikal's reemergence in 680 CE, followed by the rise of secondary powers such as Naranjo and Caracol between 700–1000 CE. Without question it was war that changed the balance of power in the region. The worst, however, was yet to come.

Imagine a scenario in which an overpopulated state with a capital of 50,000 people as at Tikal or Calakmul maxes out its long overutilized water and food resources then is hit by an extended drought. Only by turning to old alliances such as Naranjo and Caracol can they survive. This likely meant sharing military resources and thus power, especially if large-scale star wars were waged. These star wars events became the means to not only knock out your rival polity and acquire scarce resources for your state but also the means by which smaller polities could become more powerful, especially the royal families that headed them. This can be seen even at Dos Pilas and Piedras Negras, relatively small and unimportant polities before the late classic. Royal alliances and competition for urgent subsistence supplies led the lowland Maya around the Peten down a path from which they would not return intact. Warfare moved beyond seasonal fighting by elite professionals, as the archaeology and epigraphy suggests. Smaller second-, third-, and even fourth-level polities could not field military specialists as Tikal and Calakmul used to do, so they turned to commoners, the same ones who would normally be tending the fields and water supply, the river transports and *sacbees* or roads. War was waged year round, the delicate and complex infrastructure went into decline, and cities became untenable. Civilians, now armed to the hilt, fields in ruin, turned on the elite ruling classes whom they had entrusted the balance of the universe

with. The Mayan twin corn gods would lead the classic Mayan commoners to new fields in the north, from which the *teosinte* (maize) seed would sprout forth. But it was the end for classic Mayan civilization as it had been lived for over 500 years.

For many readers this scenario will seem plausible or even probable just as other explanations for the collapse have. However, since the Mayan collapse is so uneven in its nature, we need to seek an explanation that is by nature uneven. Warfare impact, as noted repeatedly in studies from across the globe, leaves an uneven mark on the landscape of civilization. Some regions, polities, and peoples may be sacked, while others realign and still others remain neutral. Just think of World War I and the differential levels of conflict evidence it left around the globe and within Europe itself. It greatly depends on where you look for war as to whether you find the evidence, but that does not mean it did not take place! Perhaps a better analogy would lie in medieval Europe or Samurai Japan with constantly shifting royal political alliances and almost constant internecine warfare. Some castles and estates saw constant conflict, some intermittent, some managed to escape. Often this was a function of a variety of factors, including geographic location, politics, and luck. Yet no one doubts the evidence that medieval Europe and Japan had a high degree of warfare at the same time as the classic Maya. The major difference is probably the prodigious written records between the regions. No doubt if no written record of the crusades into the Middle East was left, or of Portuguese voyages into Japanese trading ports, archaeologists might still be debating whether or not the events took place. That is the nature of archaeology, both good and bad.

The evidence for war as a major factor in the collapse exists in the Petexbatun and in the Peten in general. The main issue to be resolved is what the relationships were between war, ecology, and climate in the collapse and transformation of the classic Mayan world of the Peten. We might imagine a sliding scale in which war accounts for 40 percent of the destruction, ecological overuse another 20–30 percent, and perhaps climatic and other changes another 20–30 percent. However, no longer can the role of warfare as a major shaper of Mayan history be left in the background. One wonders what the ancient Maya, who at last abandoned sites at Dos Pilas or Tikal or Caracol, might have answered when asked the question of what happened to their civilization, so long successful in the Peten. Sometimes the distance of time in the research community can leave us isolated from the reality of the historical experience we seek to study. It is, after all, the perspective and worldview of the ancient ones that we are looking for.

References and Further Reading

Brown, Kathryn. *Ancient Mesoamerican Warfare*. Lanham, MD: AltaMira Press, 2003.

Chase, Arlen, and Diane Chase. "Late Classic May Political Structure, Polity Size and Warfare Arenas." In *Anatomia de una Civilizacion: Aproximaciones Interdisciplinarias a la Cultura Maya*. Edited by A. Ciudad Ruiz. Madrid, Spain: Publicaciones de la S.E.E.M 4, 1998

Connell, Samuel, and Jay Silverstein. "From Laos to Mesoamerica." In *The Archaeology of Warfare: Prehistories of Raiding and Conquest* (pp. 395–433). Edited by E. Arkush and M. Allen. Gainesville: University Press of Florida, 2006.

Demarest, Arthur. "The Vanderbilt Petexbatun Regional Archaeological Project: New Evidence and Interpretations on the Classic Maya Collapse." *Ancient Mesoamerica* 8:2 (1998), 207–228).

Demarest, Arthur, Prudence Rice, and Don Rice. "The Terminal Classic in the Maya Lowlands: Assessing Collapses, Terminations and Transformations." In *The Terminal Classic in the Maya Lowlands: Collapse, Transition, and Transformation*. Edited by A. Demarest, P. Rice, and D. Rice. Boulder: University Press of Colorado, 2005.

Diamond, Jarod. "The Last Americans: Environmental Collapse and the End of Civilization." *Harper's Magazine* (June 2003), 43–51.

Diamond, Jerod. *Collapse: How Societies Choose to Fail or Succeed*. New York: Penguin Books, 2005.

Gann, Thomas, and J. Eric Thompson. *The History of the Maya from the Earliest Times to the Present Day*. New York: C. Scribner's Son, 1931.

Gill, Richardson. *The Great Maya Droughts: Water, Life, and Death*. Albuquerque, NM: University of New Mexico Press, 2000.

Keeley, Lawrence. *War before Civilization*. London: Oxford University Press, 1997.

Marcus, Joyce. *Emblem and State in the Classic Maya Lowlands: An Epigraphic Approach to Territorial Organization*. Washington DC: Dumbarton Oaks, 1976.

Martin, Simon, and Nikolai Grube. *Chronicle of the Maya Kings and Queens*. London: Thames and Hudson, 2000.

McKillop, Heather. *The Ancient Maya: New Perspectives*. Santa Barbara, CA: ABC-CLIO, 2004.

Schele, Linda, and David Friedel. *A Forest of Kings: The Untold Story of the Ancient Maya*. New York: Morrow, 1990.

Stuart, David. "Historical Inscriptions and the Maya Collapse." In *Lowland Maya Civilization in the 8th Century A.D*. Edited by J. Sabloff and J. Henderson. Washington DC: Dumbarton Oaks, 1993.

Webster, David. "Ancient Maya Warfare." In *War and Society in the Ancient and Medieval Worlds* (pp. 333–60). Edited by K. Raaflaub and N. Rosenstein. Cambridge, MA: Harvard University Press, 1999.

Webster, David. *The Fall of the Ancient Maya: Solving the Mystery of the Maya Collapse*. London: Thames and Hudson, 2004.

Wright, Ronald. *Stolen Continents: The "New World" through Indian Eyes*. Boston: Houghton Mifflin, 1992.

4

The Chinese explorations of the 1420s reached both coasts of North and South America.

PRO   Justin Corfield
CON   Eric Cunningham

PRO

There has been intense debate over whether or not the Chinese reached America before the 16th century. Because of the vast size of the Pacific Ocean, many people had not thought it feasible to cross the ocean before the building of the caravels by the Portuguese, although the view that the Vikings in their long-boats were able to cross the Atlantic in early medieval times is now widely accepted. However, because of the size of the Pacific Ocean and its currents, as well as the very small number of islands in the North Pacific, the problems facing seafarers were enormous.

In 1947 the Norwegian seaman Thor Heyerdahl and his small crew, in the *Kon-Tiki*, proved that it was possible, in a raft made from balsa wood, to go from the west coast of South America to Polynesia. From that point, scholars started suggesting that it might have been possible for people to sail eastward across the Pacific, using the trade currents, to reach America. Heyerdahl proved that it was possible, but whether it took place was doubted.

Early Claims of Chinese "Discovery"

The idea that the Chinese might have "discovered" the Americas can be traced back to Joseph de Guignes, who suggested that Buddhist monks from China went to the Americas in the fifth century CE. These ideas have been systematically disproved, and Joseph Needham (1971) went through all the then-extent evidence for the Chinese in the Americas in the pre-Columbian period.

China is not traditionally thought of as being a strong maritime power and certainly has not been one in the past 500 years. However, in the 15th century, the Chinese admiral Cheng Ho (Zheng He in pinyin) did command a large fleet of massive ships that sailed around the South China Seas and the Indian Ocean, and as a result, it has been suggested that some ships commanded by the Chinese captains Zhou Wen, Zhou Man, Yang Qing, and Hong Bao may have visited many other parts of the world, including, possibly, the west coast of the Americas.

69

The voyages of Cheng Ho were well known to Chinese scholars, and, indeed, there are Chinese temples and communities around the South China Sea and the Indian Ocean that have long celebrated the expedition. However, these voyages were relatively unknown in the West until a former British Royal Navy submarine commander Gavin Menzies (2002) wrote his book *1421: The Year China Discovered America*, which became an international bestseller. In 2005 a large exhibition was held in Beijing to commemorate the voyages of Cheng Ho. In the book, Menzies suggests that the Chinese might have visited the Americas, and as a result, this claim has become known as the "1421 hypothesis."

Cheng Ho's Voyages

The great Chinese admiral Cheng Ho was born as Ma Ho in about 1371 in Kunming, in Yunnan Province, in the southeast of China. Cheng Ho's family claimed to have been descended from an early governor of Yunnan from the Yuan (Mongol) Dynasty, tracing their descent from the kings of Bokhara in central Asia, and Cheng Ho's father, a devout Muslim, had made the pilgrimage to Mecca. When he was 10, the Chinese loyal to the newly established Ming Dynasty captured Yunnan from the Mongols, and he was captured and sent to work in the imperial service, where he distinguished himself and became an important figure at court.

In 1400 there was a rebellion that led to Emperor Chien-wen being overthrown two years later and he was replaced by his nephew Yung-lo. Yung-lo reigned for the next 22 years and wanted to demonstrate the power of the new Ming Dynasty to the entire region. Ma Ho took the name Cheng Ho and was selected by the new emperor to sail around the "Western Oceans." The first fleet led by him in 1405 consisted of 62 ships and 27,800 men. This was a massive undertaking and extremely expensive—when Columbus sailed for the Americas in 1492, he took with him only 86 others, and various European rulers did not want to back him because of the cost of the venture. The fleet of Cheng Ho reached Champa (in modern-day central Vietnam), then sailed for Siam, Melaka (in modern-day Malaysia), to Java, and then around the western part of the Indian Ocean, reaching modern-day Sri Lanka, and returning to China in 1407.

Obviously the emperor was please with the first fleet because he ordered a second fleet to be prepared. It embarked on its voyage in 1409 and headed for modern-day Sri Lanka, where the Chinese were involved in an attack on King Alagonakkara, defeating his armies, and brought him back to the Chinese capital of Nanking as a prisoner. The third voyage started in 1411 and headed straight for India, where it then reached the Straits of Hormuz (between modern-day Iran and Oman) and returned via Sumatra. A fourth voyage was undertaken in 1413, again heading for India, and then to the modern-day Oman, with some of the fleet heading to modern-day Aden. A Chinese delegation did reach Mecca and

then traveled to Egypt, with the main fleet heading south along the east coast of Africa, reaching modern-day Mozambique. They returned in 1415.

A fifth voyage lasted from 1417 until 1419 and again reached the Persian Gulf and the east coast of Africa, with a sixth voyage, in 1421, taking back some of the foreign emissaries who had gone to Nanking on the fourth voyage. The cost of these sea voyages was enormous. As a result, with the death of Emperor Yung-lo in 1424, the imperial court decided to disband the navy and end all naval expeditions. This had as much to do with the cost of mounting the expeditions as with the intrigue in the imperial palace. However, Cheng Ho did manage a seventh voyage in 1431–1433, going to Southeast Asia, India, the Persian Gulf, the Red Sea, and the east coast of Africa. Cheng Ho died in 1435.

The details of most of the seven voyages of Cheng Ho are not contested, and details of the ships sent out and the places visited survive, both in Chinese records and in records of some of the places they visited. There are temples and plaques that remain, commemorating the voyage. Even some of the dry docks in Nanking, where the enormous ships were built, have survived. What a study of these voyages does show was that Chinese ships were easily capable of sailing around the Indian Ocean, were easily able to reach the coast of east Africa, and, if necessary, were well armed and capable of taking part in local wars. This then demonstrated that the Chinese ships of the period were capable of sailing across the Pacific Ocean, or across the Indian Ocean and around the Cape of Good Hope into the South Atlantic. The debate rests on whether or not they did so.

Illustration of the Ming Dynasty navigator Cheng Ho. Between 1405 and 1433, Cheng Ho led seven immense fleets in expeditions to Southeast Asia, India, the Persian Gulf, and the east coast of Africa. These expeditions, made up of several dozen ships carrying some 20,000 men, heightened Chinese prestige, increased trade, and added more countries to the Ming tribute system. (The British Library/StockphotoPro)

Arguments against Chinese "Discovery"

The most obvious reason suggested why the Chinese did not reach America before 1492 was that the Chinese had never claimed to have discovered the

---

A Medal for Discovering America

In 2007 a seven-centimeter diameter brass medallion was dug up in a graveyard in coastal North Carolina. That, in itself, is not astounding, but what was inscribed on the medal is. A Chinese inscription reading, "Da Ming Xuan De Wei Ci," translates as "Awarded by Xuan De of Great Ming." It refers to the reign of Emperor Xuan Zong, from 1426 to 1435, decades before Columbus sailed. It has long been asserted that Chinese explorers, most notably Cheng Ho, reached the Americas before Columbus, and this discovery, if valid, lends more credence to that theory. Medallions like this were given to those who carried the authority of the emperor, and certainly Cheng Ho would have qualified on that count. There are other theories explaining how an early 15th-century Chinese medallion ended up buried in North Carolina, but Gavin Menzies, author of *1421: The Year the Chinese Discovered America*, sees this as the first of a number of many similar findings that will reshape how Americans look at the continent's "discovery" by outsiders.

Americas, whereas they did claim to have "discovered" other parts of the world. The answer to that is obvious. Many important documents in China were destroyed in the wars and revolutions since the 1420s. The loss of archives and libraries in the emergence of the Manchu Dynasty in the 1640s, the devastation of much of the country during the Taiping Rebellion in the middle of the 19th century, and the civil war, the Japanese invasion, and the Cultural Revolution of the 20th century resulted in the loss of many histories and manuscript accounts. Many detailed histories of the period are known to have existed, but all copies appear to have been lost.

Although there are only two main written references in Norse sagas to the Vikings sailing to places that could be identified as the Americas, archaeological evidence has proven beyond any doubt that they did so. As to why the claim had not been made beforehand, Cheng Ho's expeditions certainly incited a large degree of jealousy in the imperial court, and it is easily possible that details from his later expeditions could have been deliberately hidden either by his direct enemies or by people anxious to play down the discoveries, because they might have resulted in further expeditions that would have drained the government coffers.

There are a number of surviving maps indicating that the Chinese voyages of Cheng Ho—and those of his admirals—covered a great deal more territory than had previously been thought. The Kangnido Map, dated to 1402, before Cheng Ho's first expedition, certainly shows many details of the African continent and appears to have been drawn in a stylistic fashion, possibly originally from a rough sketch. It comes from a map that was presented by the Korean ambassador to the Chinese Emperor Yung-lo in 1403, and although the original has been lost, a Japanese copy has survived. This clearly shows the Straits of Gibraltar, the North African coast, including the Atlas Mountains, and the entire

coast of the African continent, although with a large inland lake—a common perception by many explorers of the period. Because the European sailors did not cover the continent until 60 years later, the map shows that somebody in China had knowledge of the African continent before the Portuguese had sailed around the west coast of Africa. Although, obviously, it is possible to sail around the coast of Africa, following the trade currents, it seems likely that the ships going from the Indian Ocean around the Cape of Good Hope might have reached West Africa, and then possibly headed across the Atlantic to northern Brazil.

This speculation leads to the Pizzigano Map of 1424, made by a Venetian cartographer named Zuane Pizzigano. It ended up in the collection of the British antiquarian Thomas Phillips and from there, found its way to the James Ford Bell Library at the University of Minnesota. It also showed the African continent. But more important, there were four islands marked on the map in the western Atlantic. These were called Satanazes ("Devil's Island"), Antilia, Saya, and Ymana. There are no islands in the location marked on the map, but because the Europeans had problems calculating longitude, Gavin Menzies showed that it was possible that these were, in fact, Caribbean islands. He suggested that Satanazes was the modern-day island of Guadeloupe, with Antilia being Puerto Rico.

Given that Christopher Columbus did not sail into the region until 1492, the map indicates that some people reached the Caribbean at least 70 years earlier. Much more conclusively, one of the islands has a remark written on it stating that it was where volcanoes erupted, and there had been eruptions in the Caribbean during this period. It was, in fact, this map that led Gavin Menzies to begin his work, and he believes that the Pizzigano Map incorporated information from other, probably Chinese, maps. Again, as with the Kangnido Map, the Pizzigano Map does not prove that the Chinese reached any part of the Americas, but it indicates that *somebody* did go there (and return) and clearly suggests that it was possible.

The third map, the Fra Mauro Map, from 1459, has very detailed coverage of places in Asia, the Indian Ocean, and Africa, again predating the European circumnavigation of Africa. More important, it also shows an "Indian" *Zoncho*, marked as a ship, in the Atlantic Ocean. As the term "India" was loosely used in Europe to refer to Asia—leading to the native people of the Americas being called Indians, for example—and the term *zoncho* stood for a "junk," Menzies argues that this again shows evidence of Chinese ships in the Atlantic Ocean.

There are several other maps that, again, tend to point to the Chinese explorers having located parts of the Americas. The Martellus Map of 1489 includes rivers in South America identified as the Magdalena (Colombia), the Orinoco (Venezuela), the Amazon (Brazil), the Paraguay and the Parana (Paraguay), and the Colorado, Negro, and the Chubut (all in Argentina). Menzies has also managed to identify on the map the Mississippi, the Brazos, the Alabama, and the Roanoke in North America. There is also the Piri Reis Map of 1513, which,

Menzies argued, shows that Admiral Hong Bao—one of the admirals serving under Cheng Ho—was involved in charting the coast of Antarctica. Again, it shows that the Chinese ships seem to have been sailing all around the Americas, so it seems unlikely that they did not discover some part of North or South America. Furthermore, as early as 1428 the Portuguese claimed that they had a map covering the whole world, but they never claimed they had created this chart.

A more intriguing piece of evidence that the Chinese reached the Americas comes from a 1430 Chinese publication called *The Illustrated Record of Strange Countries*. It describes some animals such as the armadillo, which are unique to South America, and the mylodons, which were then found in Patagonia. Many voyages of discovery included hunting local animals and occasionally taking them back to the explorers' home country, so the interest in armadillos would certainly have been expected, given its unusual appearance.

Also, much has been made of the presence of Asiatic chickens in the Americas. The bird is flightless, and it seems likely that the Chinese fleets would have taken a supply of these for fresh meat on their voyage. They predominantly lay blue-colored eggs—both in China and also from Mexico to Chile, along the west coast of Latin America—while the European chickens lay eggs that are either cream colored or white. Ferdinand Magellan refers to finding a new breed of chicken in South America, and the references to these chickens by the Incas predated the arrival of the Spanish conquistador Francisco Pizarro. The term for the word chicken was *hualpa*, and the great Inca leader Atahualpa was called that before the arrival of Pizarro, who overthrew and killed him in 1533. Curiously, the chicken in Mexico is known locally as *tori*, while in Japanese it is *nihuatori*. The Arawaks called it *karaka*, whereas the Indians referred to the chicken as *karaknath*. Similar work has been done on the maize used in the Americas and its link to types of maize used in China, contrasting with the maize in Europe at the time.

Evidence of Chinese Presence in the Americas

To compare, once again, the possible Chinese "discovery" of the Americas with the Viking discovery of North America, apart from the Vinland Map—which is questioned by some experts—there are no other documentary sources for the Vikings' voyages to North America, except references in two sagas, mentioned earlier, and the various rune stones, some of which may be hoaxes, and others (such as the Kensington Runestone), which may be genuine. However, few people doubt that the Vikings did reach North America, mainly because of the archaeological remains of Viking settlements found at L'Anse aux Meadows in Canada and elsewhere.

As a result, researchers have been trying to locate any evidence in the Americas of sites that may have possible origin or artifacts. Menzies has identified

17 sites or documentary sightings that seem to have connections to early Chinese sailors or settlers. Two of these are on the east coast of North America. There were reports of Chinese junks off the coast of Florida by the Spanish explorer Pedro Menendez de Avilés, and when George Washington and some colleagues drained a swamp west of Norfolk, Virginia, it has been reported that they found the remains of "an ancient C Chinese sailing ship." In the Caribbean, Columbus is reported to have found a shipwreck off the coast of Guadeloupe, and Vasquez de Coronado is reported, in Hakluyt's *The Principall Navigations, Voiages, and Discoveries of the English Nation* (1589) to have seen junks with gilded sterns— "three shippes on the sea coast which bore Alcatarzes or pelicans of gold and silver in their prows." A Chinese book also cites a secret report by Columbus in which he found Chinese mirrors in "bird boats."

There are far more interesting discoveries on the west coast of the Americas. A large shipwreck was found off Vancouver Island that was longer than James Cook's *Endeavour* or Vancouver's *Discovery*. A number of Chinese artifacts, including vases, were also found nearby. An "Asian pot wreck" was found several centuries ago, and this has been linked to an unexplained group of potters who settled in the valley between Vancouver Lake and the Columbus River, operating there between about 1400 and 1700. J. G. Swan's book *The Northwest Coast* (1857) also makes mention of Chinese artifacts at Clatsop Beach, Vancouver Island, and Ming porcelain has been found on the Netarts sand pit in Oregon, along with Southeast Asian hardwood that has been dated to about 1410. This was in about the same location as some wrecked Chinese junks were illustrated in the map of California from 1543 drawn up by Juan Rodriguez Cabrillo. A Chinese brass plate was found buried at Susanville, near Sacramento, California; and there are Ming porcelain objects in local museums in southern California and northern Mexico. There also are regular reports of Native North American tribes with collections of Chinese coins, although nobody has currently dated these to any particular time period.

The English sea captain Francis Drake is reported to have "chased a junk" off the coast of modern-day California in June–July 1579 when he was circumnavigating the world in 1577–1580. Although this may refer to a Chinese ship, the reference is in passing—the vessel never posed a threat to them—and it is possible that the mention could refer to a local boat, although Drake did sail to the Moluccas and through the East Indies, where he certainly would have come across genuine Chinese "junks."

Other accounts come from John Ranking's (1827) *Historical Researches on the Conquest of Peru, Mexico, Bogota, Natchez and Talameco, in the Thirteenth Century by the Mongols*. Ranking claims in his books—he wrote another on the Mongols and Rome, published in 1826—to have spent 20 years in "Hindoostan" and Russia. Ranking studied details of the Mongol fleet, which did attack Japan, and queried whether some of the ships could have reached the Americas.

This seems far less likely than the fleet of Cheng Ho making the trip, and Ranking cites his reasons based on his studies of the items from the region, including carvings of elephants and also the copper armor in pre-Columbian America, in Peru and Mexico, as well as some characteristics in their hieroglyphics used in some parts of the region. There was, however, evidence of similarities with those used by the Mongols.

Ranking detailed wrecks of what are believed to be Chinese ships off the coast of Peru and the remains of junks found near "Taroja" in Chile. He also mentioned the presence of some animals, such as the tapir, common to both regions. Certainly, the similarities to Mongol items are interesting, but this does not rule out Cheng Ho's expeditions as being the source of these. The Yuan Dynasty ruling China until just before the birth of Cheng Ho was Mongol, and indeed the Mongols were in control of the province of Yunnan when Cheng Ho was born. As a result, it seems highly likely that Cheng Ho would have taken with him at least some Mongols, and there would be a certainty that some of the older people connected with voyages would have studied during Mongol rule, which ended in parts of China 53 years earlier, but which continued in regions up to 30 years before the actual voyage. It was not until 1402 that the Mongol rulers ceased to use the name of the dynasty.

There are numerous other similarities between items in China and those in pre-Columbian America, including fishhooks and labrets used in Ecuador that are similar to those in China. Lacquer boxes from Ming China have been found in Mexico, as have the remains of Chinese dyes. Bronzes with Chinese inscriptions have been located in Trujillo in Peru, as was a silver "idol" with a Chinese inscription, and at Nasca, also in Peru, pottery with Chinese inscriptions has been found. In addition to the early Ming brass plate that was found in California, a stone with a Chinese inscription was also found in that state. There have also been suggestions that some of the carved Mayan statues and figurines in modern-day Mexico and Guatemala have Chinese-style features, and writing on a Mayan temple wall in Yucatán appears to be Chinese in origin. The evidence of carvings of what are believed to be elephants in Central America by the Maya and by other peoples in modern-day Bolivia has also given rise to suggestions that these inscriptions or models were done by, or brought from China, as elephants had died out in prehistoric times.

Mention should also be made that in parts of northern Peru, especially in the Ancash Province, Menzies managed to identify some 95 geographic names that are words in Chinese and have no significance in any of the local dialects, such as Quechua and Aymara. Indeed, he pointed out that the name "Peru" means "white mist" in Chinese and, also the name "Ch-Li" (for modern-day Chile) means "dependent territory" in Chinese. There are also 130 geographic names in Peru corresponding to names of places in China. Indeed, the Inca roads in Peru were made using gypsum cement in a similar style to that used in China.

Another area of investigation into whether Chinese intermarried or settled in the Americas has come through analysis of DNA. In 1964 in a study of the distribution of transferrin phenotypes—proteins that transport iron in blood—Tulio Arends and M. L. Gallengo from the Instituto Venezolano de Investigaciones Cienti'ificas in Caracas, Venezuela, examined some of the people in tribes in the foothills of the Sierra de Perija in Venezuela. Some 58 percent of them had a slow-moving transferrin that was similar to that found in Chinese from the province of Kwantung, in southeast China, and nowhere else. As it has been suggested by Menzies that it is extremely likely that many of the crew for the ships of Zhou Man and Hong Bao would have come from Kwantung, and it seems highly likely that at least one Chinese person is the ancestor of some of the Venezuelan tribes people.

Recently Uncovered Evidence

The major evidence that the Chinese did discover the Americas came out after the publication of Menzies's book. A Chinese lawyer and collector, Liu Gang bought a map from a Shanghai dealer for $500 in 2001. The map was being sold at a stall in a bookshop in Fuzhou Street, Shanghai, and the inscription on it stated that the map was drawn in 1763 based on a copy of another map made in the 16th year of the Emperor Yung-lo (1418). The inscription states that the map was drawn by Mo Yi Tong and shows North America, South America, Africa, and Australia. But it does not include the British Isles. After hearing about the book by Menzies, Liu Gang revealed his possession of the map.

Known as the Zheng He Map, some skeptics believe that the map is false, while others claim that even if the map can be accurately dated to 1763, with forensic tests in Canada on the paper and the ink, we only have the copier's statement that the map was an accurate copy of a 1418 map. A forensic test of the paper showed that there was an 80 percent probability that it dated from between either 1640 and 1690 or between 1730 and 1810. Because only the paper was tested—and not the ink—there still remain questions concerning it. However, there are certainly intriguing anomalies about the map that point to it possibly being genuine. These involve Australia being wrongly placed. It is more-or-less accepted that the Chinese reached Australia before Captain James Cook. More intriguingly, California is shown as an island. It was originally thought to be such by the Portuguese and shown as one on Portuguese maps, but the Portuguese did not sail to California, making it possible that the early Portuguese maps were, in fact, copied from the Chinese map.

Conclusion

There was much debate about the research by Gavin Menzies, and some of his arguments can be demonstrated to be shaky, or to rely too heavily on other sources

that are sometimes not cited in full. Given the reported sightings, the evidence from isolated archaeological findings, and from evidence from analysis of the Indians in Venezuela, it seems highly likely that the Chinese did visit the Americas. Some scholars have sought to disprove some of the arguments by Menzies, but to prove that the Chinese did not reach the Americas, they would have to disprove all of them. It is certainly clearly possible that the Chinese had the capability to sail to the Americas—both across the Atlantic, and via the Indian Ocean and around the Cape of Good Hope, across the Pacific to the west coast.

The main evidence for the Chinese discovery of the Americas comes from the maps, and it is becoming even more obvious that some sailors had certainly visited the Americas and mapped parts of them before Christopher Columbus arrived there. This information then reached the west—Columbus refers to the sailors from Bristol, and English court records do mention an expedition to a place called "Brazil" in the reign of King Edward IV (d. 1483), with the term "Brazil" possibly referring to Newfoundland.

This might explain how Columbus knew of the existence of the Americas somewhere across the Atlantic, but the maps cited earlier show that other people had clearly known much of the western coastline of the Americas, as well as the eastern coastline, before Columbus set sail. As this predates the Europeans, there are few other options. The Japanese did not have ships capable of sailing the Pacific Ocean; their society was so insular and they had little contact with the outside world that it was unlikely such information could have reached Europe. This leaves the intense probability, rather than the certainty, that the Chinese discovered the Americas before Columbus.

References and Further Reading

Arends, Tulio and M L Gallengo. "Transferrins in Venezuelan Indians: High Frequency of a Slow-Moving Variant." *Science* 143:367 (January 24, 1964), 367–368.

"China Beat Columbus to It, Perhaps." *The Economist* (January 14, 2006): 80–81.

Duyvendak, J. J. L. *China's Discovery of Africa*. London: Arthur Probsthain, 1949.

Gilmore, Donald Y., and Linda S. McElroy, eds. *Across before Columbus?* Edgecomb, ME: NEARA Publications, 1998.

Grice, Elizabeth. "Explorer from China Who 'Beat Columbus to America'." *Daily Telegraph* (London) (March 4, 2002): 5.

Jeffreys, M. D. W. "Pre-Columbian Maize in Asia." In *Man Across the Sea: Problems of Pre-Columbian Contacts*. Edited by Carroll Riley et al. Austin: University of Texas Press, 1971.

Menzies, Gavin. *1421: The Year China Discovered America*. New York: Harper Collins, 2002.

Menzies, Gavin. *Zheng He's Voyages to Xiyang*. Singapore: Canfonian, 2004.

Needham, Joseph. "China and Pre-Columbian America." In *Science and Civilisation in China* (Vol. 4, Part 3, pp. 540–53). Cambridge: Cambridge University Press, 1971.

Ranking, John. *Historical Researches on the Conquest of Peru, Mexico, Bogota, Natchez and Talameco, in the Thirteenth Century by the Mongols*. London: Longman, Rees, Orme, Brown, and Green, 1827.

Viviano, Frank. "China's Great Armada." *National Geographic* (July 2005): 28–53.

CON

Prior to the release of Gavin Menzies's (2003) *1421: The Year China Discovered America*, the "big" problem for historians studying the Ming treasure fleets was not trying to establish proof of New World explorations, but rather trying to determine why the Ming chose not to explore any new worlds. Given the fact that its ships were large and seaworthy, its captains highly skilled, and its seafaring technologies adequate to the task of relatively long distance voyages, why, historians asked, did China "turn its back" on the sea? Why did the Ming rulers (1368–1644) decide not to send their fleets beyond the Indian Ocean and create the kind of maritime empire that less sophisticated European powers would establish a century later? If only China had seized this opportunity, historians speculated, European exploration, and the expansion of Western civilization might have been dramatically altered.

This question has been the subject of interest to China scholars throughout the 20th century. In the first half of the century, celebrated Sinologists such as J. J. L. Duyvendak and Joseph Needham (1959, 1962, 1965) wrote about the short but dramatic period in which China was the world's largest maritime power. In the past 20 years, Edward Dreyer (2007), Robert Finlay (1992, 2004), and Louise Levathes (1994) have published focused studies of the treasure fleets, offering greater insight on such matters as the motives for exploration, the diplomatic context, military history, and even naval architecture. The general consensus among historians is that the voyages were a priority only to the expansive emperor Yongle (r. 1402–1424), and were not necessarily consistent with the larger aspirations of a new dynasty trying to restore good Confucian governance after nearly a century of Mongol (1279–1368) rule. By this argument, the voyages were terminated with the death of Yongle because the cost of supporting the fleets diverted resources from more vital concerns such as restoring the agricultural infrastructure of rural villages, maintaining a good defensive posture against frontier enemies, and

rebuilding the urban commercial economy after decades of Mongol neglect. Another reason often given is that eunuch "admirals," epitomized by the great Zheng He (1371–1433), became too influential in court, prompting jealous mandarins to eliminate their rivals' opportunities to acquire prestige and power. All things considered, the question is a good one and has yet to be fully resolved. Historians over the years have written a great deal in their attempts to clarify the problem, but none has ever felt the need to challenge the validity of the base assumption that the Ming voyages were limited in scope, duration, and space, and extended no farther than the western Indian Ocean.

The entire discussion changed in 2003, with the publication of Menzies's controversial bestseller *1421*, which insists that the Ming treasure fleets discovered, charted, and even settled the American hemisphere as early as 1421, decades prior to the first Portuguese explorations in the Indian Ocean and Spanish explorations in the Atlantic. Menzies's thesis has not made the old question obsolete, but it has shifted the focus of the discussion away from its original, and arguably more substantial, historical concerns. No more do we debate the question of why the Ming fleets failed to establish a Chinese maritime empire; instead we weigh in on the problems of Menzies's methodology and take positions on the degree to which we think he has abused the practice of good history. While judgments of Menzies's project do not completely answer the interesting question of whether or not the Chinese discovered America, they do indicate that what is at stake today in the writing of history is not simply the interpretation of evidence but rather the manner in which historical questions are posed, researched, and presented. The question before us is less historical than historiographical, and to address it, we have no choice but to examine one book and its controversial author.

Menzies's thesis, bold in its position and revolutionary in its implications, has created a firestorm among historians. His book (along with its sequel *1434: The Year a Chinese Fleet Sailed to Italy and Ignited the Renaissance* [2008]) has spawned a print and Internet cottage industry of sympathetic readers who continue to furnish their own "evidence" of China's American explorations. Not surprisingly, the *1421* phenomenon has also inspired an equally vigorous community of "Menzies-bashers," primarily professional historians who accuse Menzies of poor methodology and unsubstantiated theories. Menzies's critics maintain that as an amateur historian (he is in fact a retired British submarine commander), he has not adhered to even minimum standards of professional historical research, specifically failing to consult primary sources (he does not read Chinese) and failing to provide any positive evidence for his radical claims about Chinese naval presence in the Atlantic, the Antarctic, the Americas, and Oceania between 1421 and 1423. In taking up the "con" side of this argument, it would be easy enough to say that the critics are correct, and that the case is closed. It is true that Menzies does not read Chinese; it is true that Menzies

A replica of a ship used by Ming Dynasty explorer Cheng Ho, also known as Zheng He, while under construction at a park in Nanjing, eastern China's Jiangsu Province in 2006. Cheng completed seven voyages between 1405 and 1433, leading huge flotillas of more than 100 ships and 25,000 men to 30 countries. (Sean Yong/Reuters/Corbis)

engages in dramatic conjecture, and it is true that Menzies does not—and based on the available evidence cannot—"prove" that Chinese ships arrived in the American hemisphere in the 1420s. If good history is done on the basis of clear documentary evidence, which is to say, evidence that positively proves an assertion, rather than merely raises a possibility, then there is now no conclusive historical argument to support the position that the Ming fleets discovered America.

Having said that, I would also concede that the bluntness of this ready counterargument may do some injury to the healthy practice of historical speculation, and may be too dismissive of Menzies's project, in that it fails to address completely the good questions that Menzies does raise about the existence of charts that may show regions of the Americas and the wider world prior to 1492. Had Menzies left these questions unanswered and avoided the temptation of answering them with confidence "beyond the scintilla of doubt," he may have been more enthusiastically received by academic critics, though the boldness with which he made his assertions (which gained him a large part of his notoriety) would likely have been gone as well (Menzies 2003: 127). When amateur historians bring bold, new, and unsubstantiated theories into print, becoming icons of popular culture as a result, the natural consequence is a backlash of academic ire that may be as rash in its condemnation as the contested theses are in their postulates.

China's Treasure Fleets

In July 2005, the city of Nanjing, in China's Jiangsu Province, opened the Zhenghe Treasure Ship Ruins Park, which commemorates and preserves many of the relics of the Ming Dynasty treasure fleets, which may have arrived in the Americas before Columbus. The voyages of Cheng Ho (or, Zheng He) to Southeast Asia, India, the Persian Gulf, and the Red Sea are a primary focus of the park, which includes a hall of treasures exhibiting gifts Cheng Ho received during his voyages. The centerpiece of the exhibit is a 200-foot replica of a treasure fleet ship, although Cheng Ho's flagship would have been over twice as large. His fleet consisted of some 300 ships and 37,000 sailors. The fleet began in 1405, when the first Ming Emperor, Yongle, commissioned Cheng Ho to sail in order to establish trading partners and demonstrate China's prowess as a naval power. It is noteworthy to point out, however, that although the Chinese fleet voyaged as far as any European explorer, they did not build forts to establish a permanent presence, preferring instead to amass a fortune through trade.

Menzies, whose *1421* project resembles such fictions as the *Da Vinci Code* or *National Treasure*, may not deserve to be called a serious historian, but he probably does not deserve to be called a charlatan either. He is certainly not the first history buff to be mistaken for a professional historian by an enthusiastic public.

The controversial theory of the Chinese discovery of America is, despite some embarrassing holes, an important topic, not only because of what the evidence may or may not tell us, but because the controversy itself reveals some of the limitations of "good" historical scholarship and invites us to consider what the correct role of the historian should be. It also forces us to consider why unsubstantiated theories can so quickly and permanently work their way into the consciousness of a public looking for truth in history. That a society groping for satisfaction in a troubled world will often embrace pseudo-history as real history says a great deal about the vitality of that society's historical aspirations, to say nothing of its confidence in its own historical assumptions. This section, in opposing the thesis that the Chinese discovered America, will not only address the argument and its evidence, but will also consider some of the larger questions about history, historiography, and historical consciousness that it presents.

What the Evidence Shows

Thanks to reliable court documents and biographical records of the Ming Dynasty, we know that the third emperor Yongle commissioned a series of voyages to the "western ocean." He did this, apparently, to "display his shoulders in strange lands in order to make manifest the wealth and power" of his empire

(Dryer 2007). Yongle's goal was not merely to project Ming power abroad, but also to draw the maritime nations of Southeast Asia and Africa into tributary relationships with the Chinese court. The voyages, beginning in 1405, marked something of a departure from early Ming foreign policy. Yongle's father (and dynasty founder) emperor Hongwu (r. 1368–1398), chose a path of diplomatic isolation after years of fierce campaigns to expel the Mongols from China. In the early years of his reign, he banned trade with foreign countries, set to work on reconstructing the Great Wall, and repudiated the expansionist policies of the Mongols, who had ruled China for nearly a century. Yongle, on the other hand, was ruthlessly self-aggrandizing, determined to increase his own glory and the prestige of his empire (Tsai 2001). After a bloody four-year rebellion against his nephew, the Emperor Jianwen, Yongle seized the throne in 1402 and began a multifaceted project to expand Ming power over land and sea. By 1410, he had reopened hostilities with the Mongols, made incursions into Vietnam, and launched the fabled treasure fleets—all, as Edward Dreyer observes, in defiance of the wishes of the mandarins, the Confucian scholar-officials who staffed the imperial bureaucracy. In the final analysis, Yongle's dreams of expansion came to little, as these adventures were too costly to sustain. After Yongle's death in 1424, his imperial successors and the Confucian bureaucracy took decisive steps to curtail the military and maritime adventures that had cost China so much in terms of lost lives and wasted treasure. While the decommissioning of the treasure fleets was easy to justify from an economic standpoint, it is clear that court politics played a significant role in these decisions. The core of the political controversy was the long-standing suspicion of the mandarins, who saw themselves as the rightful administrators of the realm, toward the court eunuchs, who enjoyed great influence as palace guards, generals, and, for a while, sea captains. The life of the legendary "admiral," Zheng He, who commanded these voyages, gives evidence of how power relationships outside the traditional channels of Confucian scholarship and government were treated by the mandarins as a threat to political stability.

Often portrayed as a great sailor in the legendary maritime tradition of Islam, Zheng He's chief qualification for the position of commander-in-chief of the treasure fleets was not his expertise as a skipper, but rather his unwavering loyalty to Yongle. As a young boy, Zheng He had been taken prisoner during the Ming wars against the Mongols. He was castrated and forced into service in Yongle's household while the latter was still an imperial prince and a commander in his father's army. As Zheng He grew up, he became a great soldier, serving with conspicuous bravery during his patron's quest for the throne. While it is true that Zheng He came from a devout Muslim family, he was neither a mariner by training nor an experienced explorer prior to being placed in charge of the treasure fleets (Dreyer 2007).

Six voyages were launched between 1405 and 1422, with a seventh and final voyage commissioned in 1431, seven years after Yongle's death. By all

accounts, the size and composition of the treasure fleets was staggering, especially in comparison to the three 85-foot caravels that Columbus took to the Caribbean in 1492. As an illustration, the first Ming fleet got under way with 62 treasure ships, each over 400 feet in length and 180 feet in the beam. These large flat-bottomed ships displaced over 20,000 tons; each was furnished with nine masts, sealed watertight compartments, fore and aft anchors, and sternpost rudders. They were accompanied by 255 support ships carrying water, horses, gifts for overseas Ming vassals, and a complement of about 27,000 sailors, ambassadors, and men-at-arms. The seven fleets were sent to destinations in Vietnam and Southeast Asia, India and Sri Lanka, the Straits of Malacca, the Gulf of Aden, the Arabian Peninsula, the Red Sea, and eastern Africa as far south as Mozambique. As noted above, these voyages were chiefly diplomatic missions, designed to show the grandeur of the Ming emperor, to establish tributary relations with the rulers of the Indian Ocean area, and to collect medicinal and zoological specimens. Because these goals seem to stand in such stark contrast to the later goals of Spanish and Portuguese explorers (which might be characterized by the old, but not inaccurate, mnemonic of "God, gold, and glory"), they have often been interpreted as "peaceful" missions. It is difficult, as Dreyer points out, to consider these deployments of enormous ships filled with soldiers as peace missions, especially in view of the fact that Ming soldiers engaged in hostilities with locals on several occasions, but this interpretation has been used by Chinese and non-Chinese historians alike to suggest that China's geopolitical ambitions, also in contrast to the West, have always been, and still are, benign (Dreyer 2007).

The crux of the debate is the sixth voyage of 1421–1422, which Menzies claims is the voyage that brought the treasure fleets to the western lands. Conventional accounts tell us that the purpose of this voyage was to return long exiled diplomats to their home countries. While there is evidence to show that the fleet separated into smaller task forces, there is little else to support the idea that these subordinate forces sailed into the Atlantic or Pacific Ocean. We know that the voyages, with one exception, came to an end with the death of Yongle in 1424. The mandarins, who had long opposed the treasure fleets and military expansion, convinced the new emperor, Hongxi, to suspend the voyages, and the government took forceful steps to prevent any resumption of maritime activity. Needham and Levathes both contend that deck logs, sailing directions, and ship's plans were destroyed by Confucian officials in later decades of the Ming Dynasty, in the hopes of making sure that the empire would never again waste "myriads" of treasure and grain on such non-Confucian priorities as seafaring (Levathes 1959). Exactly what records the officials destroyed, and precisely why they did so, is a matter of debate (Dreyer 2007). It is beyond doubt that the Ming Dynasty turned away from maritime activity, but proving that only the records of Zheng He's voyages to America were destroyed—while making

use of the records of the first five voyages that were kept—seems to be an uphill battle. The argument in favor of the Chinese discovery of America absolutely depends on the belief that key documents about the new world were destroyed, because the gap in the historical record is the only basis available to the creative historian who wishes to supply the missing narrative. With that, let us turn to the Menzies argument.

The Menzies Thesis: Strategies and Assertions

It is not the purpose of this section to present an item-by-item refutation of Gavin Menzies's argument, but because Menzies is the author of the only existing thesis that asserts that the Chinese discovered the new world, we have to grapple first with the argument he makes and the evidence he uses to support it. When we do so, we find that the argument and the evidence are two sides of one rhetorical strategy to offer conjecture as conclusion, and to offer random and often irrelevant data as if they were a preponderance of proof. Menzies makes no serious attempt to prove that the Ming fleets discovered America—he just asserts that they did, and makes this assertion the ground of proof for the evidence that follows. His methodology relies on three rhetorical strategies: (1) making the absence of a historical record the essential proof of his thesis, (2) treating detailed discussions of peripheral matters as if they were central issues, (3) offering personal intuition as if it were professional expertise. The combined effect of these strategies is to lull an uncritical reader into thinking that the conclusion, because it is stated so boldly, must be true, and that the author, because he demonstrates a certain degree of real knowledge on such technical matters as naval architecture and celestial navigation, must also be a legitimate authority on history. The speculations he makes, then, strike the reader as the kind of courageous new arguments that institutional scholars are afraid to make, rather than merely the fantasies of an amateur historian who desperately wishes they were true. If one accepts Menzies's conclusion as premise, then the book seems plausible, especially in the latter (and in absolute terms, *least* convincing) chapters, which strike readers in much the way that any book rich with data might—as an almost mind-numbing avalanche of supporting evidence that leaves them overwhelmed with the feeling that the author must be right. When we look more closely, though, we can see that the "conclusion as premise" method has no substance, and we have to admit that his evidence, as interesting and overwhelming as it is, and as cleverly deployed as it may be at times, really proves nothing.

In relation to the first strategy, *1421* is constructed over a void. The starting point of Menzies's argument is the claim that the mandarins destroyed the real evidence of the far western voyages. As we have seen, the officials after Yongle did bring an end to the treasure fleets, and they did make efforts to suppress future seafaring, efforts that included the destruction of some documents associated with

Zheng He's voyages. Nevertheless, it is impossible to know the content of destroyed evidence, and neither Menzies nor anybody can state that the destroyed material was categorical proof of voyages to the Western Hemisphere. In other words, the evidence that makes the book "work" is missing evidence—it is not the case that the Chinese discovery of America has been long known or widely suspected and that Menzies has heroically found it—it is rather the case that no evidence for the Chinese discovery of America, as such, exists at all (Menzies 2002: 81–82; Needham 1959: 558). Without these critical pieces of evidence, all we have are conjecture, circumstantial evidence, and intuition. This is precisely where Menzies's second and third strategies come into play. The evidence of carved stones, artifacts from shipwrecks, observation towers, "Chinese" maps, transplanted animals, and human DNA scattered throughout the world is only persuasive if we accept Menzies's thesis upfront. We are asked to do so for two reasons: first, there is no *other* cohesive explanation for all of this evidence, and second, because Menzies's experience as a Royal Navy submarine commander gives him the ability to confirm the explanation as true on the basis of his experience and trained instinct. At one point he writes: "If I was able [*sic*] to state with confidence the course a Chinese fleet had taken, it was because the surviving maps and charts, and my own knowledge of the winds, currents and sea conditions they faced told me the route *as surely as if there had been a written record of it*" (Menzies 2003: 81, emphasis added). All told then, we have an original thesis, passionately stated, illustrated by a wide variety of data, none of which adds up to historical certainty. The opening chapters of the book contain a large quantity of good, common knowledge and a small number of controversial proposals. By the end of the fourth chapter, "Rounding the Cape," this ratio has been inverted, with pure speculation as the ground for every argument, and bits of evidence put forth as the unifying hub of the speculation. Let us look at the how the book goes from being "interesting enough" in the introductory chapters to simply implausible by only the fourth chapter.

The Menzies Narrative: Construction and Shortcomings

In the introduction to *1421*, Menzies tells us that his research was inspired by the chance discovery of a nautical chart from the year 1424 depicting islands that seemed like they might be in the Caribbean. Knowing that no Europeans had sailed to the Caribbean prior to 1492, Menzies wondered how knowledge of these islands, if they really were in the Caribbean, could have become known to Zuane Pizzigano, the Venetian who made the chart. After comparing the relative strengths of the maritime nations of the mid-15th century, Menzies concluded that only China had the resources, the technology, and the seafaring expertise to launch the kind of explorations that would have brought mariners into contact with such far-flung regions, and he decided that he would find additional evidence to support this conclusion.

The first chapter gives us a portrait of a resurgent China, enthralled by the expansionistic vision of the emperor Zhu Di (Yongle). Without dealing concretely with the historical context of Ming China's diplomatic situation, Menzies asserts that Yongle was driven by the desire to achieve world domination on land and sea. To this end, he launched the monumental sixth voyage of 1421. The rich depictions of Chinese economic and intellectual life lend an air of authority to Menzies's narrative, although as the endnotes show, this information was gleaned from the most general sources available and have little to do with the specific event of launching the sixth fleet. In the second chapter, historical events that took place over 50 years apart are conflated as the single event that led the mandarins to ban maritime voyages forever. The first event was a catastrophic fire in 1421 that damaged the capital, Beijing; the second was the 1477 destruction of the records of Zheng He's voyages by the mandarin Liu Daxia. By placing these events in such close narrative proximity, Menzies gives us the impression that the imperial government had acted swiftly to end the treasure fleets, not only to appease the anger of heaven, but to restore common sense to the realm even while Yongle was still alive. Thus, the most complicated aspect of the historiography of the entire debate is glossed over in anecdote and confusing chronology. In the third chapter, Menzies makes his first reference to the problem of determining longitude—an almost irrelevant concern to the primary content of the chapter, which deals with shipboard life and routine. The information on navigation seems almost a distraction to the main story, but it lends a measure of credibility to Menzies, whose subsequent narrative recounts details on shipboard life, complete with the habits of embarked concubines, all from secondary sources. This chapter ends with the division of the fleet into four squadrons, each commanded by a different eunuch admiral: Hong Bao, Zhou Man, Zhou Wen, and Yang Qing. It is almost impossible to figure out why he is certain these men were Zheng He's subordinate commanders because none but Zhou Man is identified as such in any prior study. Yet, coming as it does at the end of a long catalog of reasonably factual material, the casual reader is disinclined to challenge the veracity of the assertion or any of the grandiose speculations that follow.

In the fourth chapter, which discusses the rounding of the cape of Africa, the book's logic begins to break down, although it is here that Menzies most confidently asserts his authority: "[A]s I began to trace the voyages of the great treasure fleets in the 'missing years' from 1421 to 1423, I was entering familiar territory, making use of knowledge and skills I had acquired over many years' experience as a navigator and commanding officer on the high seas" (Menzies 2003: 81). In this chapter, a set of carved stones that claim Zheng He and his companions visited 3,000 countries is presented as the primary evidence for the tales of exploration that make up the rest of the book. To supplement this thin foundation, Menzies again invokes his navy experience, offering his knowledge of the "winds, currents and navigational problems the Chinese admirals encountered"

as a guarantee that they had rounded the cape and turned northward into the Atlantic (Menzies 2003: 83). Statements of historical fact that proceed only from claims of intuition cannot be taken seriously by academic historians, yet the rest of the arguments in *1421* are built upon just these kinds of claims. Of particular importance to the thesis are Menzies's instincts concerning the role played by the Venetian explorer Niccolo da Conti, who as a young man converted to Islam and sailed to India and the Middle East. Menzies not only "felt certain" that da Conti met the Chinese chronicler Ma Huan in India, but that he also brought Chinese charts back to Europe, unlocking the mystery of how Europeans had knowledge of "Chinese discoveries" prior to 1492 (Menzies 2003: 86). Whether da Conti did any of these things can only be speculated, yet by the end of the chapter, Menzies has already used these speculations to make a concrete conclusion: that Chinese ships sailed south out of the Indian Ocean.

At the risk of sounding too dismissive, there is no need to go into any details concerning the rest of the book. The remaining chapters have no real evidence and no new methodological strategies—only more cases of looking for evidence to support premade conclusions, and more statements such as "The conclusion is inescapable: Chinese fleets must have brought chickens to South America" (Menzies 2003: 126). The fog of data and the boldness of the claims eventually bring more confusion than clarity to the reader's mind. By the end of *1421*, the Ming treasure fleets have not only beaten all sensible odds by rounding the cape of Africa, they have also discovered North and South America, Antarctica, Australia, and the North Pole—yet somehow completely missing Europe and West Africa in the process! If Menzies had added an epilogue claiming that the treasure fleets were actually made up of submarines, and he knew this because he himself had commanded submarines, it would not appear too fantastic in light of all the other claims.

I should probably add, at the risk of being facetious, that the best argument against the Chinese discovery of America is the impressive body of evidence to support the theory that the Spanish and Portuguese *did* discover America. Leaving aside the matter of whether or not the fully operational and thriving civilizations of America were really "discovered" by anybody, as well as the fact that any theory that the Chinese discovered the "new" world is just as dismissive to pre-Columbian civilization as its European counterpart, the Europeans obviously found something in America worth conquering and keeping. The argument that Chinese mariners arrived and then left, never to return, could be taken as a slight against their ability to recognize the value of opportunities that sustained contact with a new hemisphere may have afforded. It is one thing to say, as the conventional arguments do, that China's cultural priorities kept it from ever considering exploring the new world. It is quite another to say that once having seen the American continents, they simply abandoned them. Menzies's argument is held up as a kind of compliment to China, but on closer examination it looks like an

insult—it argues that the Western exploring nations, as puny as they were, possessed such power as to be able to accidentally find, deliberately conquer, and then utterly despoil an entire world, while China lacked even the ability to realize what it had stumbled upon. Do we imagine that a Chinese state so obtuse could have prevented the spread of Western imperial expansion? Historians and students of history alike need to be very vigilant in watching for the concealed implications of even "flattering" historical arguments.

Conclusion and Historiographical Considerations

Rather than begin with a conservative hypothesis and seek to demonstrate it with substantial facts, Menzies proceeds as if his theory were irrefutable fact, and takes every bit of evidence, pertinent or not, as "proof." Not only then does Menzies put the "cart" of his conclusion before the "horse" of meaningful data, he also used the assumed fact of his conclusion as a justification for stuffing a wide range of additional historical mysteries into his ready-made thesis. So, without proving that the Chinese discovered America, he conveniently uses the "fact" that China did discover America to solve a variety of problems involving premodern cartography, natural history, anthropology, and zoology. What we have then is less a case of historical research than a creative exercise in using random facts as plausible evidence to support a hypothesis. This practice is so common in popular journalism, television, and historical fiction that it is easy to see why a curious and "truth-seeking" public would accept it. A society accustomed to such alternate histories as *Raiders of the Lost Ark, The X-Files, The Da Vinci Code*, and *National Treasure*—which is to say, a society that has lost its credulity in official knowledge and simply *knows* that there is more "out there"—is easy prey for such a strategy, and "serious" history, because of its built-in limitations and sometimes stifling assumptions, has a difficult time convincing the public not to buy into sensation and hype. Arguably, serious history could do a better job appealing to people who, by nature, are not necessarily looking for truths so small that they *can* be proven beyond the shadow of a doubt by clear documents.

It is often said that there is no such thing as objective, value-free history. All histories, even those that purport to be scientific and objective, are serving some larger cultural purpose than merely telling a true tale of the past, even if that purpose is merely validating the thing we have come to call "objectivity." When scientific historians claim that their works are true because they are based on primary source documents—authentic statements of law, policy, philosophy; journals, letters, and speeches—they do not necessarily concede that even these primary sources are conditioned by the emotions, attitudes, and social contexts of their own producers. A primary source may be the most authentic source, but it is by no means the most accurate rendering of absolute historical truth, because the writing of history will invariably support the self-defined interests

of the historian and his or her culture. In a sense then, history writing does as much to create and reinforce consciousness as it does to discover objective truth. It remains to be seen what future historians of China will do with the *1421* theory of Chinese exploration. If China should depart from its Maoist revolutionary heritage and seek a new identity as a modern liberal nation, the Menzies thesis may serve to add legitimacy to Chinese claims for dominance in a progress-determined world. On the other hand, should China decide to redefine itself as a post-Enlightenment or postmodern global power, it may seek to distance itself from any suggestion that it neglected to create the historical conditions that only Europe could bring to maturity. It is a matter of some concern that the coming decades may see any number of new appropriations of the Menzies thesis, despite the fact that the thesis itself fails to prove its assertions.

As both interpreters and makers of reality, historians are always engaged in a battle for certainty, and the stakes are always high. Accordingly, historians should be conservative in generating their hypotheses and very strict in interpreting their data. The debate over whether Chinese fleets "discovered America" before the Europeans is indeed an issue that has the potential to reshape world historical consciousness, but at present, there is no historically sound reason to believe that it happened.

References and Further Reading

Dreyer, Edward L. *Zheng He: China and the Oceans in the Early Ming Dynasty, 1405–1433*. New York: Pearson Longman, 2007.

Finlay, Robert. "How Not to (Re)Write World History: Gavin Menzies and the Chinese Discovery of America." *Journal of World History* 15, 2 (June 2004): 229–242.

Finlay, Robert. "Portuguese and Chinese Maritime Imperialism: Camoes's Lusiads and Luo Maodeng's Voyage of the San Bao Eunuch." *Comparative Studies in Society and History* 34, 2 (April 1992): 225–41.

Furnish, Timothy. "Is Gavin Menzies Right or Wrong?" History News Network. March 10, 2003. Available at http://hnn.us/articles/1308.html (accessed June 9, 2010).

Hsu, Mei-Ling. "Chinese Marine Cartography: Sea Charts of Pre-Modern China." *Imago Mundi* 40 (1988): 96–112.

Levathes, Louise. *When China Ruled the Seas: The Treasure Fleet of the Dragon Throne, 1405–1433*. New York: Oxford University Press, 1994.

Menzies, Gavin. *1421: The Year China Discovered America*. New York: Morrow, 2003.

Menzies, Gavin. *1434: The Year a Chinese Fleet Sailed to Italy and Ignited the Renaissance*. New York: Morrow, 2008.

Needham, Joseph. *Science and Civilization in China*, Vol. 3. Cambridge: Cambridge University Press, 1959.

Needham, Joseph. *Science and Civilization in China*, Vol. 4, Part 1. Cambridge: Cambridge University Press, 1962.

Needham, Joseph. *Science and Civilization in China*, Vol. 4. Part 2. Cambridge: Cambridge University Press, 1965.

Prescott, Victor. "1421 and All That Junk." The 1421 Myth Exposed. Available at www.1421exposed.com/html/1421_and_all_that_junk.html (accessed June 9, 2010).

Tsai, Henry Shi-shan, *Perpetual Happiness: The Ming Emperor Yongle*. Seattle: University of Washington Press, 2001.

5

The technologies that allowed Europe to dominate the world were all imported from the East: compass, lateen-rigged sail, gunpowder, windmill, stirrup, movable type.

	PRO	David Blanks
	CON	Talaat Shehata

PRO

The year 1500 is often seen as a turning point in the history of our planet. It is a convenient date. It marks what some call the Age of Discovery, when Europeans learned the art of ocean navigation that enabled them both to reach the New World and to open a direct sea route to India.

According to the 18th-century economist Adam Smith, "The discovery of America, and that of the passage to the East Indies by the Cape of Good Hope, [were] the two greatest events recorded in the history of mankind." (1776). A hundred years later Karl Marx agreed, as did the sociologist Max Weber. For many intellectuals, the Age of Discovery brought an end to the Middle Ages (500–1500 CE) and the beginning of modernity, the period when the West came to dominate the globe.

Yet the path to power was neither quick nor easy. If we were to take a survey of the world in 1500, we would see that at the time when Europe is commonly believed to have been launching itself toward greatness, it was actually quite backward and underdeveloped compared to the rich, populous, sophisticated empires of Asia. In today's terms, it would have been seen as a third world country—rural, poor, and illiterate. The great cities and great civilizations were farther east. If asked to predict which people were poised for domination, an informed 15th- or 16th-century observer would have said Ming China, or perhaps Mogul India, or Ottoman Turkey, or Safavid Persia. No one would have bet on Europe, a marginalized region far distant from the centers of wealth and imperial might. But our informed observer would have been wrong. Five hundred years later Beijing, New Delhi, Istanbul, and Tehran are filled with teenagers sporting jeans and T-shirts, eating at international coffee shops, using PlayStation, and watching first-run Hollywood movies. Europe, and the West in general, have taken over the globe, not only militarily, but economically and culturally as well.

93

Given the relative weakness of the West in 1500, how did this happen? What made it possible? What changed? This is *the* question in world history today. How was it, to paraphrase Marshall Hodgson (1974), that European merchants were able to outproduce, outtravel, and outsell anyone, that their physicians were better able to heal than others, that their scientists put the rest to shame, that their armies, their politicians, their businesspeople, and their performers are the ones whose influence resonates throughout the world and not those of some other culture?

Any number of explanations have been suggested: political organization; military tactics and experience; the emergence of comparatively free and vibrant civil and commercial societies; population growth, bureaucratic expertise, urbanism, capitalism; the nation–state; and the Industrial Revolution each, at one time or another, has been put forth as the leading cause of European success.

Today, however, the two factors that are most frequently mentioned are technology and biology. One school of thought has it that technologies imported from the East, such as the compass, the lateen sail, gunpowder, and movable type, were instrumental in enabling Europeans to establish themselves as a global power; the other, that it was actually European resistance to animal-borne diseases that enabled them to conquer rival civilizations.

The Technology Factor

The first theory is connected with what is sometimes called the "military superiority thesis." It argues that Europe went through a "military revolution" in the 16th and 17th centuries, made possible, in large part, by technologies borrowed from China, India, and the Arab world. These technologies, most importantly the use of gunpowder, were then improved upon, and the resulting weapons, it is said, combined with their vast practical experience in warfare, gave Western nations a decisive advantage. This seems like a satisfying answer, especially to people living in the 21st century. We can easily see the relationship between war and power, and we are accustomed to thinking of global dominance in terms of military might.

Other historians have modified the military superiority thesis by suggesting that advanced military technology and strategic superiority alone were not enough. They speak not merely of a "military revolution" but of a "modern revolution." They argue that there were a whole series of inventions and innovations that led to the "Rise of the West." Ploughs, windmills, water mills, paper, printing, books, advances in mining and metallurgy, new ship designs, sails, navigational devices like the astrolabe and compass, the spinning wheel, and the clock all were brought to Europe from elsewhere, and all contributed, in one way or another, to Europe's ability to dominate the globe. This, too, seems sensible, especially, once again, in the modern world, where the link between technology and economic and political ascendance seems self-evident.

Was this really the case? Or are we, with hindsight, projecting modern assumptions about how the world works back onto the past? The connection between technology and power is clear today, but did it also work that way prior to the Industrial Revolution of the 19th century? Let's take some examples of the transfer of technologies from East to West and see if this was in fact what led to European success.

As explained in the introduction, compared to most of Asia, the Latin West was an underdeveloped region in the Middle Ages (500–1500). As a civilization it had a lot of catching up to do. Government was ineffective; security lax; standards of living low; and cultural achievements minimal. It would be several hundred years before order could be established even within the bounds of Europe—and several hundred more before Western peoples would reach demographic, economic, and military parity with the vast and powerful empires of the East. Furthermore, as suggested above, without the network of cultural diffusion that stretched from Mongolia to the Mediterranean, Westerners might never have managed to move beyond their own borders onto the world stage.

To understand how this process worked, let's begin with a comparative analysis between East and West. This will allow us to see where things stood *prior* to the impact of technology and biology. As a starting date, let's choose a familiar one, namely, 1066, the year of the Battle of Hastings, one of the most decisive military encounters in European history. When Harold Hardrada rushed to the coast to defend his kingdom from the invading Normans, he had with him about 8,000 Saxon infantrymen. William the Conquer had some cavalry, but the total size of his force was about the same: fewer than 20,000 men in a legendary battle that shaped the entire future of the country. Yet at the same time, on the other side of the world, in a place the English had never even heard of, the emperors of Song China (960–1279) were regularly fielding armies that exceeded 1 million. According to the Domesday Book, a medieval census, this was roughly equivalent to the entire population of England at the time. Indeed, in the Middle Ages, even the largest Latin Christian armies maxed out at 30,000 men. It is a striking contrast.

Moreover, in terms of military technology, whereas William the Conqueror's army at least had archers and mounted knights, Harold's didn't even have horses. Imagine what would have happened if they had faced the Song. Armed with clubs, spears, swords, axes, and rocks, the Saxons would have been devastated by even a tiny Chinese division of mounted archers fighting with sophisticated crossbows and small incendiary bombs. For that matter, had any of the Latin Christian kingdoms been located in East Asia, they quickly would have been crushed and absorbed into the Chinese Empire.

It is even more striking to learn that these advanced weapons were supplied by government factories that also made thousands of suits of armor and millions of iron arrow heads annually. That the Chinese government was able to supply

---

The Largest Invasion Fleet in History:
Kublai Khan's Plan to Invade Japan

In 1281 the Mongolian Emperor Kublai Khan launched a bold plan: to invade Japan with over 4,400 ships. However, neither the ships nor the 70,000 men they carried succeeded in conquering the island nation. Having built the ships in the short period of one year, the mere existence of the fleet was amazing. However, the scant evidence left in Japanese scrolls and more recent underwater archaeological studies show that the Mongol fleet met a similar fate as did the miniscule (in comparison) Spanish Armada. Off the coast of Japan's Takashima Island, Japanese archaeologists have uncovered Mongol anchors, helmets, swords, and arrows in great numbers. From the evidence found, scholars have argued that the Mongols were unsuccessful in defeating the Japanese samurai that defended the islands and had to remain on their ships. From the direction of the anchors of the ships found, it appears that some type of large wind event, possibly even a typhoon, did the fleet in. Archaeologists who have analyzed the ship remains maintain that the fact that the fleet was constructed in such haste meant that many of the ships were merely riverboats that were much more poorly constructed than normal Chinese ships, which were among the best in the world. Given that fact, it's no surprise that a typhoon could have been enough to sink the fleet.

---

its vast army through organizing production in state-controlled factory systems shows the extent of their technical superiority. Sophisticated blast furnaces, fueled for the first time by coke, were capable of producing huge quantities of iron and steel. In addition to weapons, the iron was used to produce coins. Already by this time more than half the state's revenue was being collected in cash, another sign of the advanced nature of Chinese civilization; and paper money had been in circulation since the early ninth century. It would not make its appearance in the West until the first regular bank notes were issued in Sweden in 1661.

The Chinese moved their supplies and raw materials to the factories in the capital via a complex shipping and transport network that depended upon the construction of an extensive system of canals and locks. Some of these ships were armored, driven by treadmills and paddle wheels, and equipped with projectile-throwing machines. Their navies, like the armies, were huge. In 1281, in an attempted invasion of Japan, the Chinese fleet consisted of an estimated 4,400 ships. Like the Spanish Armada that attempted to invade England in 1588, the attack was destroyed in large part as the result of storms; but the Spanish fleet was only a fraction of the size, comprising a mere 130 ships.

Yet another indication of the superiority of Asian power and technology was the series of voyages made between 1405 and 1433 by the Chinese Admiral Zheng He. The first of these seven two-year cruises, to Calicut on the west

coast of India, consisted of more than 300 ships that carried 27,000 sailors and soldiers—a force that might well have been able to invade and take England where the Spanish had failed. Sixty-seven of these vessels were what the Chinese called "treasure ships," between 370 and 440 feet in length and 150 to 180 feet wide with up to nine masts, some 90 feet high, and displacements of between 2,000 and 3,000 tons. By comparison, the European Age of Discovery, hailed by Smith, Marx, and Weber as a turning point in the history of the world, was a rather paltry affair. When Columbus "discovered" the New World, he had but three ships with him; and when Vasco da Gama made his famous voyage to India in 1497, he had but four. The largest flagship was a three-mast ship, less than 100 feet in length, that displaced a maximum of 300 tons. Columbus's three caravels were even smaller.

Even by the late 18th century—a time by which it is commonly assumed Europe was in a position of dominance—the British, who by then had become the principal Western trading nation and the foremost European sea power, were still unable to make inroads into the Chinese market, enjoying but a modest presence in the single port city reserved for the use of foreigners, Guangzhou. And they were nowhere near capable of defeating the Chinese military. (That would have to wait until the Opium Wars of 1839–1842.)

Thus when, in 1793, the British sent a mission to China to request more extensive exchange privileges, the Emperor Qianlong sent back a startling letter to King George III. In it he thanked the British for what he saw as their submission to the Celestial Empire, and, although he excused their ignorance due to the fact that they lived on a remote and isolated island far from the center of civilization, he nonetheless declined what he considered an audacious request on the part of "barbarian merchants." There was nothing the Chinese wanted or needed from England and nothing the English could do about it.

Nonetheless, by the late 18th century, the West had advanced considerably: the English as well as the Dutch, French, Spanish, and Portuguese had in fact established forts and colonies throughout the world even if they could not be said by that time to "dominate" the globe. In the post-1500 era, as international trade and contacts increased, the merchants, diplomats, sailors, and learned men of Latin Christendom had gradually become aware of the vast economic and technological gap that separated them from the more developed societies of the East; and accordingly, little by little, they had begun to borrow Asian ideas and adapt them to their own needs and purposes. These initiatives, carried out on an individual, case-by-case basis were, in turn, integrated, bit-by-bit, into a culture that slowly came both to value—and to give public space to—entrepreneurship and innovation. And it was these initiatives that eventually gave the West its competitive advantage.

The process of borrowing and improving upon technologies, often military in nature, began well back in the early Middle Ages, as early as the eighth century,

when the stirrup, which had spread from China eastward through the Mongol, Indian, Persian, Arab, and Byzantine worlds, reached the West sometime around the early eighth century. Like the various technologies that followed, the barbarian horseman of the remote West improved upon the method until they were able to develop a type of military technology that would at last allow them to hold off Arab and Slav invaders.

The compass was another early adaptation. Invented by the Chinese around 960 CE, it was known in the West by 1180 CE at the latest, when the Englishman Alexander Neckam described it in his book *De Utensilibus*. In those days it was a rather crude device consisting of a magnetized needle floating in a bowl of water, but by the early 13th century, Indian captains were using a more sophisticated version to navigate. And by the 15th century, it was common throughout the Mediterranean. Like the history of many technologies, that of the compass is typical in that through the records we can see the device being mentioned with increasing frequency and gradually making its way from East to West, but it is difficult to know when it was first used, who made the innovations, where, or how it was transmitted. It is important for our story, however, because the compass helped make deep water, year-round shipping possible, and without it, Europeans would not have been able to sail around the world.

The stories of paper and printing are similar to those of the stirrup and the compass. Paper was invented during the Han Dynasty (220 BCE–9 CE). Although there was other progress as well, including advances in iron and especially silk production, an achievement that brought the distant Chinese culture to the attention of the Romans, arguably the most far-reaching invention was paper, which was first made by Han craftsmen from hemp and tree bark. In Asia paper money, paper shoes, wallpaper, and even toilet paper came into use centuries before the technology was transferred to the West. Block printing, too, was known as early as the seventh or eighth century. The diffusion of these ideas through Tibet, India, central Asia, and the Islamic world eventually brought paper to Europe via Spain in the 12th century. And although it is unclear whether Guttenberg was or was not influenced by Chinese techniques when he developed a Western version of movable type in 1455, one thing is certain: Europeans used this technology as a means to further their conquests. Because he was able to read reports of Cortés's expedition against the Aztecs, Pizarro had some idea of what to expect in Peru. Books allowed Pizarro to know his enemy, how he lived, how he thought, and how he might react. This knowledge may well have given him an edge with the Inca. Whether it did or not, what we know for certain is that in 1533 several hundred Spaniards armed with guns and horses brought down the entire Incan Empire.

The conquistadors and other settlers would not have reached the Americas at all if it had not been for important new developments in ship design. The first ships capable of crossing the Atlantic emerged as hybrids from the dockyards

This bronze cannon is believed to be the oldest yet discovered; it is dated to 1332 CE. The Chinese invention of gunpowder is associated with the experiments of Daoist alchemists in the early centuries CE, but its use in such weapons as mines and flame throwers did not begin until the 10th or 11th century. (Instructional Resources Corporation)

of Portugal and Spain; and the influences that went into their creation came from as far away as the North Sea and the Indian Ocean. With two or three masts, caravels were square rigged when running before the north-east trades on the voyage out, but they also sported lateen sails, borrowed from the Arabs, that enabled them to move in and out of coastal estuaries and to maneuver in shifting winds. Other features, too, such as the sternpost rudder, were likewise adopted from the Arabs, who in turn may have gotten it from the Chinese. Hulls also underwent significant modification. Collectively, these additions and modifications resulted in ships capable of circumnavigating the globe.

Mounting cannon in them made them capable of much more. Gunpowder, which was invented in China around 900 CE, is central to the military superiority thesis. As mentioned above, by the 11th century Song warriors were using both bombs and fire lances to fight of the Qidan, an invading Mongolian people from the North. Within a century, the fire lance—a sputtering burning weapon intended more to frighten than to kill—was transformed into the first crude gun. Craftsmen did this by placing the ignition hole at the rear of the cylinder, which provided for a single discharge of lead pellets and pottery fragments. William of Rubrick, papal envoy to the Great Khan, was in China in those days, and not long after his return to Europe, Rubrick's friend, Roger Bacon, published the first Latin account of gunpowder and fireworks. Perhaps this was a coincidence, but somehow the

idea of gunpowder spread from China through Russia to the West and there, at the hands of generations of technicians, scientists, and soldiers, gun weaponry was refined to the point that Europe soon became the leading arms manufacturer in the world, a distinction it maintained until the 20th century.

In the Age of Discovery, however, it was cannon that made the difference. Already by the 14th century a model had been developed that was a sort of bottle-shaped canister that discharged arrows. Later, stone balls were tried; and eventually iron ones; and then founders learned to cast cannon that could withstand the stress of larger charges. Much of the progress was a result of an arms race among the nations of Europe. There the competitive nature of small states and a burgeoning capitalism created the environment for investment and innovation in weapons technology. In China, where commerce and manufacture were under imperial control, the incentives were lacking for experimentation; and ironically, by the late 15th century, for a variety of reasons, the Chinese government began to turn landward, away from the sea. It was in those same years that European captains began installing their newfound cannon on their newly designed ships, making them an unstoppable naval force. For centuries the Arabs, Persians, Indians, and Chinese had been engaged in massive, sophisticated, and largely peaceful international trade centered on the Indian Ocean. Shortly after 1500, the Portuguese, followed by the Dutch and the English, came blasting their way in.

The Biology Factor

These heavily armed incursions into Eastern trade, the long distances traveled, and the barbaric nature of Western society have led historians to call the Europeans the "Mongols of the seas." This is true both culturally and technologically. As Arnold Pacey (1991), among others, has argued, advances in weaponry and ship design were to the Europeans in the 15th and 16th centuries what the stirrup and bow had been to the Mongols in the 13th century.

But biology, too, played a role, especially in the New World and those regions such as South Africa, Australia, and the Pacific Islands that were remote from Europe yet similar to it in climate and geography. Thus in addition to the "military superiority thesis" as a means of explaining Western domination, there has also arisen what might be called the "biological superiority thesis." William McNeill put it in an interesting way. According to him, Western power was a function of two things: macroparasites and microparasites. Macroparasites were the people who spread across the globe like an endemic disease, infecting and taking over other civilizations; microparasites were the germs they carried with them, the germs that helped them win. There is strong evidence in support of this idea.

Even at the time of the first Spanish conquests, which took place on the Canary Islands near the West African coast in the 15th century, missionaries like Friars Alonso de Espinosa and Abreu de Galindo readily admitted that they

would not have conquered the native Guanches if it had not been for the devastating effects of germs. "The mortality was so great," wrote Espinosa in *The Origin and Miracles of the Image of our Holy Lady of Candelaria* (1594), "that the island remained almost without inhabitants, they having previously numbered 15,000. . . . If it had not been for the pestilence, it would have taken much longer, the people being warlike, stubborn and wary."

Like the technologies the Europeans used in their conquests, the diseases they brought with them were also from Asia. The history of epidemics is complicated and still, to an extent, speculative, but ultimately it comes down to the question of evolution. In brief, like all organisms (or like technologies for that matter), germs continuously change and develop new forms. Epidemic disease is caused by germs that have been given the chance to evolve among social animals in crowded conditions. This means that a dense animal population is required for the spread of diseases like rinderpest, a deadly cattle plague, or, more recently, bird flu—or Simian immunodeficiency virus, which is found in West African chimpanzees and is the source of HIV (human immunodeficiency virus).

When dense animal populations correspond with dense human populations, which was often the case in settled and growing Eurasian agricultural communities in the premodern world, these pathogens are sometimes transmitted to humans, where they mutate into purely human diseases. Measles, for example, come from rinderpest. So too with the other great killers, such as measles, tuberculosis, and smallpox which came from cattle or flu, which came from pigs and ducks.

These were the diseases that were the most devastating for the indigenous peoples of the Americas and elsewhere. In the past, they were devastating for Europeans and Asians as well, but over time, the inhabitants of Eurasia developed immunities, so that the descendants of those who survived became less likely to die from infection. This was not the case for isolated populations in other parts of the world. Thus, when Spanish sailors arrived on shores that were, in effect, virgin territory for diseases that had become commonplace in Europe, they introduced germs that wiped out entire populations of people who had no natural immunities.

In Mexico, as on the Canary Islands, germs fought for the conquistadors. When Cortés first reached Tenochtitlan with his tiny force of less than 200 men in 1519, they were beaten back with heavy casualties, for like the Guanches of the Canary Islands, the Aztecs were stubborn and warlike. But when Cortés and the remainder of his men returned to Tenochtitlan early the following year, nearly half the Aztecs had died from smallpox, including the emperor. By 1618 it is estimated that Mexico's initial population of around 20 million had been reduced to less than 2 million. The sad story of Pizarro and the Incas is much the same. By the time he and his men made their way to the Inca capital, Cuzco, the emperor, Athuapulca, and many of his people had already been killed by the biological advance guard, making it far easier for the Spaniards to defeat the survivors with

their horses, guns, and swords. The same thing happened in North America. All told, it is estimated that European diseases killed up to 95 percent of the Native American population during the 16th and 17th centuries.

Similar tragedies took place elsewhere. Having first run aground near Cape Town in 1647, the Dutch eventually built fortresses and trading posts, and the first case of smallpox among the native Khoisan was reported in 1713 with predictable results. In the late 18th century, the Englishman Captain Cook "discovered" Australia, bringing with him syphilis, gonorrhea, tuberculosis, and the flu, along with typhoid and, inevitably, smallpox and measles. As elsewhere, mortality rates ranged from 50 to, in some cases, 100 percent, when entire tribes were wiped out by previously unknown epidemics.

Conclusion

Our task here has been to determine whether it was technology or biology that enabled the West to exert a global influence, but as with all complex historical questions, it is difficult to come down firmly in favor of a single, overwhelming cause. Circumstances change according to time and place, and there are many factors in addition to technology and biology that helped determine the outcome.

One such influence, nearly impossible to measure, is what we might call "geographic luck." Europe was fortunate. Although long isolated from the centers of world power and commerce, in the early modern period it became a hub of international economic activity, perfectly positioned to take maximum advantage both of its inheritance from the East and of its newfound wealth in the far West. Thus, for example, the fortunate discovery of silver in Peru allowed Europeans to exploit their technological and biological legacy and to provide themselves with the cash needed to penetrate the markets of the Indian Ocean. And it was lucky, too, that as the industrial revolution began to heat up in England, factory owners found themselves sitting on seemingly inexhaustible deposits of coal. Although unplanned, in the long run, Europe's geographic position gave it a decided advantage in the power struggles of the 16th, 17th, and 18th centuries.

A second factor, equally difficult to measure, was the influence of ideas. Simple and rather obvious discoveries, such as clever navigational techniques and a knowledge of the winds of the south Atlantic, had clear benefits, but what of political and cultural forms? How can we determine their impact? Because Europe failed to become a single, unified empire (as was the norm in Asia), it gave rise to a dangerous military and political instability that forced states to modernize or perish. This atmosphere of competition permeated the business community as well, which began to reward innovative and aggressive commercial activities, thereby giving rise to a spirit of individualism and entrepreneurship. Independent thinking and experimentation increased among scientists and scholars, too. Combined, these intangible qualities enabled Europeans to exploit

opportunities and resources in ways the world had never seen. Overseas conquests were made possible not only by technological and biological advantages, but also by the ability to organize and to govern, to tax peoples and extract profits from nature, to marshal productive forces, and to be able to adapt to changing circumstances through the creative use of reason.

A third mitigating factor in this debate, one that bears some reflection, is the question of what properly constitutes a technology. The cases of military and nonmilitary inventions listed above are obvious enough, but what about a practice or procedure, such as the three-field system, which allowed medieval farmers to put more land into production every growing season? Is that, too, a technology? Or money, which fueled economic growth, was this an invention or an innovation? Remember: it's not just the physical coin or piece of paper, but the whole idea, with all its attendant institutions, such as banks and government bureaucracies, that allowed international trade and finance to flourish.

What about the introduction of new species? The conquistadores could not have defeated the Aztec and Inca Empires without their sophisticated weapons; equally, however, they could not have sustained a large European population without their plants and animals. But do these also count as "technologies"? In the strictest sense, perhaps not; but for our purposes we should perhaps define technology in the broadest possible terms. Ships and guns and (sometimes) germs made their initial victories possible, but sustainable political and cultural domination would not have been possible without all the other techniques, practices, products, and procedures that the Europeans brought with them.

In fact it was the entire suite of Western technologies, including books, maps, charts, and mining techniques, along with various types of mills and ploughs and other labor-saving devices, that created a competitive advantage. The same observation can be made in biological terms. It was not just the germs they carried with them, but also their crops and animals—wheat and sugarcane and later cotton; and horses and cattle and chickens and pigs—that helped Western settlers occupy and control new continents and islands.

And yet, even when these additional factors are kept in mind, there *is* an answer to our question as it is here framed, for when geography and chronology are taken into consideration, that is, when we think about where and when Western power manifested itself, there seems little doubt that ultimately it was through the use of new technologies (originally Eastern and mostly military) that the rise of the West was made possible.

If we are talking about *world* domination, then it is clear that biology was only an advantage in areas that were isolated from Eurasian populations. The impact of European germs on their real rivals, the Arabs, Turks, Persians, Indians, and Chinese, was minimal. These were urban/agricultural civilizations that had built up immunities in the same way the Europeans had. It should likewise be noted that elsewhere, such as in sub-Saharan Africa and parts of Southeast Asia,

germs actually worked against the invaders. Malaria, yellow fever, dengue fever, dysentery, and all manner of tropical parasites effectively prevented Europeans from penetrating the interiors of these regions until the development of yet more new technologies—this time medical—in the 19th century. Only then was Europe able to conquer and to colonize. This suggests that even without biological superiority in places like the Americas, South Africa, the Pacific Islands, and Australia, Europeans eventually would have prevailed through other means. Germs were useful in some places, but useless or positively deadly in others; thus it is something of an overstatement to say that resistance to animal-borne disease allowed the Europeans to dominate the globe. Not to mention the fact that germs would not have had any impact at all had the Europeans not developed the means to transport smallpox and influenza to these isolated regions in the first place.

Likewise, when we consider the timing of all this, it becomes clear that what we have mostly been talking about so far is European success in the pre-modern era, but success in setting up colonies, forts, and trading posts is not the equivalent of conquering the world militarily, economically, and culturally. Even in the late 18th century, as we saw in the example of England and the Chinese market, the world's greatest wealth, power, and population still remained in the East, where the European presence was a paltry thing. On the eve of the Industrial Revolution they hung on in the Indonesian archipelago but with the sufferance of local potentates, and most of Africa remained virtually untouched. True, Adam Smith saw the discovery of the Americas and the route to India as decisive in world history, but he also understood that China was far richer than the regions of Europe, which leads us to ask: What it is we really mean by "dominance"? There is no question that Europeans began to make their power felt throughout the world in the period between 1500 and 1800, but it was not until after the Industrial Revolution in the late 18th and early 19th centuries—a technological revolution all the way—that Europeans and European culture and society could truly be said to have come to rule the world. And in this, the impact of disease on biologically isolated peoples such as the Amerindians and aborigines played no role whatsoever.

References and Further Reading

Buck, David D. "Was It Pluck or Luck that Made the West Grow Rich?" *Journal of World History* 10 (1999): 413–30.

Crosby, Alfred W. *Ecological Imperialism*, 2nd ed. New York: Cambridge University Press, 1986.

Diamond, Jared. *Guns, Germs and Steel*. London: Vintage, 1997.

Hodgson, Marshall G. S. "The Great Western Transmutation." In *The Venture of Islam* (Vol. 3, Book 6, Chap. 1, pp. 176–200). Chicago: University of Chicago Press, 1974.

McNeill, William H. *Plagues and Peoples*. Garden City, NY: Anchor Press/ Doubleday, 1976.

McNeill, William H. *The Pursuit of Power*. Chicago: University of Chicago Press, 1982.

Mokyr, Joel. *The Lever of Riches*. New York: Oxford University Press, 1990.

Pacey, Arnold. *Technology in World Civilization*. Cambridge, MA: MIT Press, 1991.

Parry, J. H. *The Age of Reconnaissance*. Cleveland: World Publishing, 1963.

Smith, Adam. *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Dublin: Whitestone, 1776.

Thompson, William R. "The Military Superiority Thesis and the Ascendancy of Western Eurasia in the World System." *Journal of World History* 10 (1999): 143–78.

CON

The basic question to be answered in this section is what actually provided the impetus for European domination of the world? Dominance from immunity to animal-borne and transmitted diseases might have played an important role; but, clearly, if one were to carefully thumb through the available historical, archaeological, and anthropological evidence, he or she would discover that many other important factors were as important, and in some cases (i.e., migration, displacement, and never-ending psychological stress and despair) other factors were of equal importance. For all concerned, conquerors and conquered, European domination didn't come or recede easily. Over the past 500 years, domination has been a historical phenomenon, which has, and continues to be a reality with which most, if not all, nations and small and tribal societies, on a global basis, have had to or are now contending with in their daily policies and affairs. The vibrant secularism that it brought has interestingly ingratiated itself into the age-old conflict since ancient days, between the forces of traditionalism and the forces of modernity, in most national and tribal histories. Therefore, the "dominance from immunity" may, instead of only being a medical historical fact, also include the immunity gradually acquired from the metaphysical, philosophical, legal, political, often racist, social, and cultural historical constructs that have evolved over the centuries within secularism and modernity's often very fluid and pliable nature.

Historical and archaeological evidence has established that within a century after the arrival of Christopher Columbus on present San Salvador's shores, 95 percent of the Native American populations had been decimated by disease. Diseases such as smallpox, diphtheria, measles, cholera, yellow fever, malaria, influenza, and whooping cough, to mention only a few, ravaged the ranks, religious, political, and social tribal structures of the indigenous inhabitants. It was

devastating and horrific. As the renowned scholars Alfred W. Crosby (1972) and William McNeill (1976) elaborated in their works, this had to occur as a direct result of the contact of alien cultures for the first time in their separate histories. It was inevitable. Their arguments, which have had loyal supporters over the past three decades, insisted on the fact that the indigenous populations were "virgin" populations. They were virgin in the sense that their geography and demography were established prior to their contact with the Europeans. They believed that the epidemics introduced by the Europeans penetrated this "virgin soil," in that the Native American inhabitants didn't have the immunological readiness to fend off these multiple varying strains of diseases, which the Europeans had acquired immunity from as a direct result of having survived most of them during their individual and collective childhoods. This argument, for the next three decades, would be a significant one that influenced the writings of many well-intentioned scholars, not only in the ever-fascinating field of American studies, but also in the equally vibrant fields of African history and world history. But, sadly, much of this reasoning has had its subtle and often very subconscious racist constructs to it, from which one could logically draw that from the late 17th, 18th, and 19th centuries, racist constructs had already been established by briefly dominant groups of European powers, which attempted to justify its fleeting global dominance in pseudo-scientific and crude racist assumptions.

Through this, racism's insidious nature forced it to justify its presence by attempting to cook the biological facts available to most, if not all, academic practitioners. By doing so, classified variations between the conquered people and other races were created, all, of course, for the "greater" good. Categories, subgroups, and subgroups to the subgroups were diligently created. Also, with the arrival of the 1980s, a new form of cultural racism replaced the old; this new racism based its premise on essentialist and cultural grounds. Much more subtle than the earlier 17th, 18th, and 19th centuries variety, it manifested itself by placing much more emphasis on the "differences" in cultural heritage and traditions, with emphasis on the distinct nature of each of the groups being studied, measured, and discussed.

In that context, one needs to pursue the impact of colonialism and nonepidemiological constructs in the decline of Native American and other indigenous groups and in their encounters with European settlers and forces. The impact of epidemic disease on these native populations was no doubt huge and catastrophic. Yet, if one were neutral in assessing what the comprehensive reasons were as to the demise and disappearance of these populations, one would need to take into consideration other primary reasons that contributed significantly to their lack of regeneration. That would have to include the impact of relocation on these diverse groups, the crushing phenomena of slavery, and, according to the deceased but yet well-regarded and profound American historian William Appleman Williams (1988), the urge and desire of the Puritans to be left alone to their own devices,

---

Sir Jeffrey Amherst and Germ Warfare

Some scholars have argued that during Pontiac's Rebellion, which followed the French and Indian War, while the Odawa war chief Pontiac was laying siege to Fort Pitt, the commander of the fort, General Jeffrey Amherst, gave the Indians blankets from the fort's smallpox isolation ward in an attempt to deliberately do what nature had been doing for over a century: decimate the Indian population. Letters between Amherst and Colonel Henry Bouquet seem to bear out the fact that Amherst was not averse to using such a tactic:

Bouquet to Amherst, June 23, 1763
Captain Ecuyer writes me that Fort Pitt is in good state of defence against all attempts from Savages, who are daily firing upon the Fort; unluckily the Small Pox has broke out in the garrison, for which he has built an Hospital under the Draw Bridge to prevent the Spreading of that distemper.

Amherst to Bouquet, July 7, 1763
Could it not be contrived to Send the Small Pox among those Disaffected Tribes of Indians? We must, on this occasion, Use Every Strategem in our power to Reduce them.

Bouquet to Amherst, July 13, 1763
I will try to inoculate the [Indians] with Some Blankets that may fall in their Hands, and take Care not to get the disease myself.
As it is pity to expose good men against them I wish we would make use of the Spanish Method to hunt them with English Dogs, supported by Rangers and Some Light Horse, who I think effectualy extirpate or remove that Vermin.

Amherst to Bouquet, July 16, 1763
You will Do well to try to Innoculate the Indians, by means of Blankets, as well as to Try Every other Method, that can Serve to Extirpate this Execrable Race.—I should be very glad [if] your Scheme for Hunting them down by Dogs could take Effect; but England is at too great a Distance to think that at present.

Source: Papers of Jeffrey Amherst, Library of Congress.

---

realized by exterminating those who stood in their path, as the Puritan fathers, mothers, children, grandchildren, and great-grandchildren, expanded their illusory or purported "isolationist" American Dream out to the western frontier.

This, among other factors, such as the messianic nature of past and current evolution of "liberal democratic" traditions, which evolved in the Western Hemisphere, especially in United States and North American continent, under the crucible and dire conditions of Western expansionism (which in time, would translate into global economic, financial, and military expansion), would thereby completely negate the earlier dominant argument throughout the first half of the

20th century by another deceased and well-known American historian Fredrick Jackson Turner, in his "Frontier Thesis," which attempted to claim that what determined the unique character of the political and social experiment in the United States was its westward expansion—where hypothetically, it had reached its ultimate destination once it had reached the Pacific Coast. All future developments within the continental United States would, in accordance with Turner's thesis, then supposedly fold onto itself. The premise for Turner's argument seemed very short-sighted to William Appleman Williams, since his strong assertion was that this political, economic, financial, and social "expansionist" approach undertaken by American policymakers throughout much of the 19th century was only getting started once it had reached the Pacific Coast. The United States and other old European powers, if they could keep up with the United States' growing and expanding potential and power at the end of the late 19th and early 20th centuries, only sought continued outward (meaning global) territories. The quest for empire, therefore, according to Williams, "was the only way to honor avarice and morality. The only way to be good and wealthy" (1988). According to Williams, this "unique Frontier" phenomena that Turner, along with many future American historians that he had mentored or influenced with his argument, so eloquently sought to elaborate and propagate missed the totality of the forest from the few trees they had mistakenly imparted too much significance to. The western frontier expansionism of Turner would, once it had reached the Pacific Coast, morph into the global American expansionist expectations that would reformulate future American policymakers thoughts. First, the Spanish American War of 1898, against the declining Spanish Empire, would provide the United States with not only a foothold into Central and South America but also with the acquisition of the Philippines, which it would give to the United States as a foothold into Pacific and Asian affairs. This would allow it, according to Williams, to continue the charade of being an "honest" broker with its Open Door policy to China, Indochina, and the region, while old European powers attempted to continue to bolster their influence and presence in the continent. According to Williams, it was very disingenuous. Its global expansionism, beginning with Admiral Perry's "opening," at gunship point, of Japan's self-imposed isolation of over 2,000 years in July 1853 and later involvement in China's internal affairs was a direct outgrowth of that early stage of American western frontier expansion. The motives were purely avarice and market oriented in its continuing expansionism. This, in time, would allow all future political, diplomatic, economic, financial, religious (primarily the evangelical variety), social, and cultural American global policies to revolve around that core nexus, in the late 19th, throughout the 20th, and into the 21st centuries, and beyond. According to Williams, the entire enterprise was, and continues to be, coldly calculated and implemented. With continuing inward and outward territorial conquest, market expansion, and war, Williams was of the view that in time, it became the easily

deferred to and convenient manner by which all future American domestic and foreign policy could be viewed through an executable prism, which continued to be defined and redefined by its elites and policymakers. Therefore, according to Williams, disease and displacement in American and European policies were allowed to perform their deadly "magic," and their global policies toward the native populations would be that of nonintervention in domestic native affairs or brazen policies of benign neglect. Once the native populations ranks had been decimated, as Williams strongly stated, as if the powers-that-be had used "bombs over Hiroshima and Nagasaki," then, according to him, gunfire would be used to "remove the hardy." All this, therefore, evolved as a natural outgrowth of the earlier nightmares one often associated with the decimation within the Native American ranks, from the late 15th to the late 19th centuries, which were caused by many of these widespread diseases and epidemics. Puritan isolationist and progressive policies required a first few steppingstones to embark on its initial virgin voyage and forward into history. It began, but according to Williams, clearly didn't end with the Native Americans. Future history would inform us.

As recent scares regarding the H1N1 or "swine flu" virus have demonstrated to a large swathe of the global community, epidemics are not merely a far distant past phenomenon. Even though, by many historians' assessments, between 47 and 54 million Native Americans perished within the first century of European arrival on their shores in 1492, the Great Influenza pandemic of 1918–1919 killed more people than any epidemic prior or since. The estimate is that in one year, on a global scale, over 40 million people lost their lives. More people died in that pandemic in one year than during the four years of the bubonic plague, or Black Death, which struck Europe and the Mediterranean world, when between a fifth of the global population and a quarter of the American population had become infected. AIDS would contribute to an emphatic exclamation point in the past two decades. This recent growing pandemic might just claim in a few short decades two to three times as many victims as the Great Influenza pandemic had at the beginning of the 20th century; making it, unfortunately, the worst pandemic in human history. Other potential epidemics that we might find ourselves having to contend with during the 21st and 22nd centuries are man-made varieties such as within a bioterrorist setting. This, if not intelligently and persistently confronted and defeated, could far surpass any of the figures established for the Great Influenza of 1918–1919 and the current AIDS epidemic.

Crosby and McNeill's "virgin soil" argument would unintentionally have contained within its core constructs a mid- to late-19th-century racist "scientific realist" construct often used by the dominant European powers' academic and policymaking elite groups during their global hegemonic influence of that period. The idea of the "noble savage," which had also grown pervasive during much of the 18th and 19th centurie's, would subconsciously inform Crosby and McNeill's "virgin soil" arguments in that late 20th century. It was all undertaken

for the "greater" good. In "scientific realism," its theories were promulgated so that through illegitimate means it would legitimize white European supremacy over all conquered people throughout the world. The biological facts available to most academic practitioners and later policymakers, who would seek to administratively implement them, were meticulously schemed. Categories, subgroups, and subgroups to the subgroups were diligently created. They created and classified hair types, skin types, skull shapes, blood types, eye types, nose types, and on and on. In time, variations were not only limited to those hair-splitting details, but they began to cast negative dispersions on the intellectual and moral basis and skills of the subordinate minorities within the European powers borders, but, more specifically, to the conquered natives within the borders of their extensive empires. With that historical backdrop, Crosby and McNeill unintentionally and inadvertently found themselves attempting to scale that slipperiest of slopes. For to attempt to convey the idea as they did, that the depopulation of the Americas was a direct outgrowth of the contact between disease-immune Europeans and a "virgin" native population in the Americas, along with other native global populations that were temporarily conquered by the European powers, was not giving complete justice to the other equally important factors— whether it be displacement, migration, malnutrition, psychological stress, exhaustion, or calculated policymaking initiatives, to mention a few. Crosby elaborated: "Virgin soil epidemics are those in which the populations at risk have had no previous contact with the diseases that strike them and are therefore immunologically almost defenseless" (1976). This argument erroneously places the "population at risk," or the "innocents" for our argument, in the position of the passive agent who erroneously had never encountered any other sort of disease or epidemic in their diverse encounters, over thousands of years, with other native tribes and populations, prior to the arrival of the Europeans. According to that argument, we are supposed to suspend our rational reasoning abilities and allow for time to stop, and then permit it to start with the arrival of the European powers. On its merits, within a historical context, this seems quite ridiculous.

The "noble savage" of the 18th and 19th centuries was defined as an "innocent" and humane side of man being untouched by the grimy nature of human industrialization and evolving civilization, living in a blissful state of "goodness." This, as we can see, morphed into Crosby and McNeill's disease and epidemic argument, that the "innocents" of the New World and the global "innocent" native populations were also living in a state of blissful "goodness," only to have the nasty, dirty, disease-carrying, and biologically immune Europeans, as they would have their readers believe, come into these multiple states of Eden and let loose their vile disease-infested baggage on the poor and completely defenseless community of "innocents." The argument evolved during the earliest phases of capitalism, from within the 18th and 19th centurie's multiple narratives, which continued in organic form into the 20th and 21st centuries. Those narratives

comprised a fertile anthropological and historical imagination as to the "true" conditions and "nature" of the primitive.

The rise and impact of the Romanticist movement in American and European literature also was strongly felt. All these "new" ideas with regard to the "noble savage and primitive" had continued to saturate and influence future readers and scholars, whether accurately or inaccurately, in how Western men and women need to perceive non-Westerners people, and especially the subjugated indigenous populations in their perceived "care." Therefore, in regard to this argument, as has been demonstrated, there's much more that informs the reasons for the destruction and subjugation of the native populations in the Americas, and throughout the globe, besides the Europeans and the Africans that they brought with them being immune from childhood and natural selection of many of the diseases and epidemics that helped decimate native populations throughout the world. Thankfully, this is still considered among many scholars and historians as a critical and important field of study that continues to deserve the gaze of many practitioners within the field and would provide more informed and less superficial answers to some of the very significant questions being posed.

The European immunity that we might need to consider is instead of a merely medical historical variety, which has been demonstrated to play an important but not totally singular role and what later evolved into a more
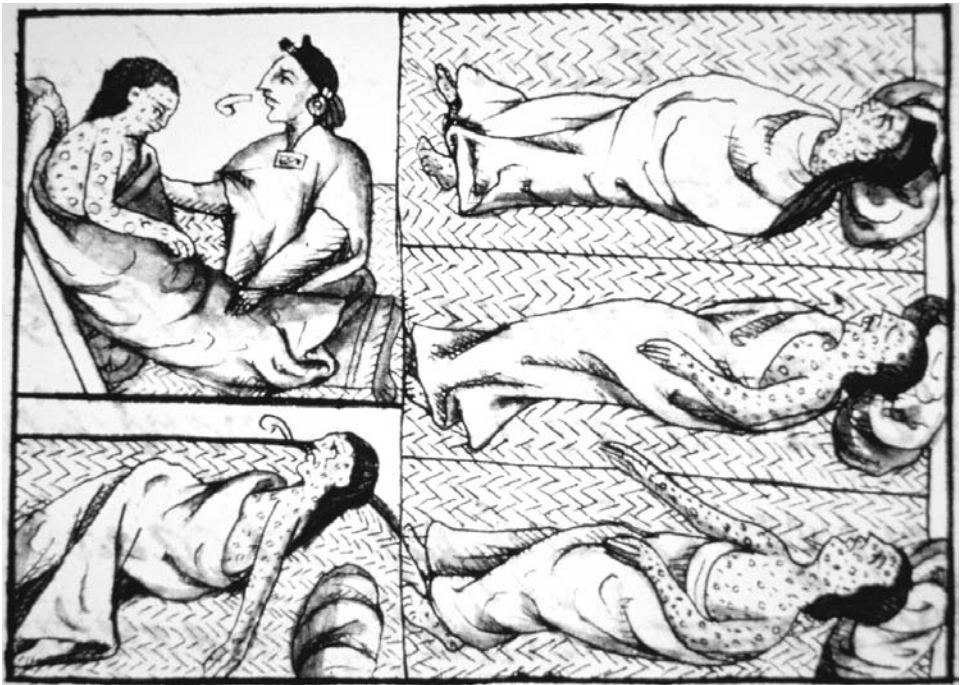
Aztecs dying from smallpox introduced by the Spaniards, from the *Florentine Codex*, about 1540. (Peter Newark American Pictures/The Bridgeman Art Library)

institutional metaphysical, philosophical, political, legal, religious, social, and cultural sort. The evolving modernism of the European powers would significantly impact nations and small and tribal societies. It would pit the historical forces of traditionalism against modernity in ways that many of the native populations had not been prepared to do. Migration, growing ethnocentrism, race-mixing, and the conceptual underpinnings of how one interpreted, or was even allowed to interpret, what "freedom" was would influence their acceptance of them. This played an important role in the evolving identities of most of these diverse tribal groups within the indigenous populations. Most of the migrations in the New World by the native populations, not necessarily being voluntary after their contact with Europeans as the Spanish, Dutch, English, French, and Portuguese, as they had often been voluntary through their earlier seasonal and climatic changes prior to their contact with the Europeans, were directly influenced by the labor needs of the European settlers and their political and trade representatives. The labor systems that were established skirted around the concept of freedom, especially in Central and South America, and brought to most of their shores coerced migrants to work the fields and gold and silver mines. This also explains the more blatant form of institutionalized and coerced migratory labor practices imposed by most of these European powers, between the 16th and 19th centuries, during the rise and eventual ultimate collapse of the transatlantic slave trade. But, most scholars and historians would agree that pure financial incentives should not explain the conditions of the native population and resettled slaves from Africa and other regions of the world. According to these peoples, there were religious and cultural factors that needed to be understood, which could better inform us as to the growth of the abolition movement in Britain, Europe, and the United States and its ultimate positive impact on the institutionalized status of the transatlantic trade by bringing it to a gradual end.

This central quest for a reformulated identity by the native populations, in the attempt to cope with the chaotic intrusions imposed upon their individual, family, and tribal lives, consistently challenged them in finding a recognizable and secure "fit" between the traditionalism that they had grown comfortable with over the ages and the unceasing modernity introduced and often coercively imposed upon them by their European neighbors and foreign taskmasters. It was William Appleman Williams's searing imagery of the messianic need for limitless Western expansion, driven by an equally potent strand of messianic economic and financial zeal for domestic and outward territories, that the evolving industrial, technological, and military forces of modernity were put at the European powers' and their settlers' regular and daily disposal to "remove the hardy." Modernity had made the Europeans immune.

With that in mind, the Europeans could then proceed ever so efficiently in dismantling all the traditionalist constructs of their conquered populations. The native populations were perceived as occupants of a preindustrial, precommercial, and

premodern space, within the larger growing, dynamic, and avarice-focused industrial, commercial, and modern society. The natives were noncapitalist; the Europeans were capitalists. The natives needed to be marginalized, if not eradicated. The process of eradication was so that despite their marginalization, the European powers always feared that they could pose as pockets of resistance to the larger outward expanding commercial, industrial, and technical machinery that they sought to create and propagate. But, on the other hand, if "tamed," then the Europeans might be able to continue exploiting them for their labor, their lands, and resources and keep them indefinitely, if not permanently, in an underdeveloped state. For the Europeans, it was sweet, a classic win-win case. The indigenous peoples rarely, if ever, digressed from their lose-lose conditions and status.

At the height of European religious missionary zeal, indigenous populations were converted to Christianity. Then, with the passage of time, secular ideas were introduced to the natives, informing them that their newly acquired religious convictions needed to be reined in, since the Constitution had asserted the right for all to be free from religious teachings and rule; the government had no legal rights to impose any specific religion on any tribe, state, or nation. This secularism also emphasized that all political actions by any varying groups and people within the nation were to be based on reason, facts, and evidence instead of religious dogma, belief, or faith. The simpler ancestral times, with each passing day, seemed to most of them like a distant disappearing mist on the horizon. But, aside from the potential resistance that they might have posed to the aggressive onslaught of European modernization, many members of the native population adopted the European religion(s) imposed upon them in order for them to assimilate. This, of course, does not deny the fact that a few natives had earlier converted out of true and deep conviction in the new faith(s) they had adopted. But, as a whole, historical evidence indicates that the earliest native converts had done so primarily as an instinct to survive the daily and regular onslaughts by the European settlers that surrounded them. But, the irony was that the more they sought to fend off the Europeans by adopting or assimilating with their religious and cultural rules and practices, the more they were mistrusted and alienated from the larger European body. In time, and well into the 20th and 21st centuries, they would pay dearly for their readiness to let bygones be bygones. They would suffer the highest unemployment, alcoholism, and drug rates in the United States; in some cases over 50 to 60 percent. Facilities to meet their health care needs were, are, and continue to be meager. Their youth see high preteen pregnancy rates, inferior educational programs, and dilapidated schools in their miserable economic circumstances. Over 55 percent of their male youths are in the larger society's prison systems. Good mental care, once released from jail, was meager to nonexistent. This contributed to a large rate of recidivism within their communities. Suicide rates have and continue to be regularly off the charts. They experience a high mortality rate and are often the victims of violent

crimes. Many of these young males do not live into their mid-20s; if they're really fortunate, they might reach their 30th birthday.

This set of evolving conditions clearly allows us to better understand an important escape mechanism employed by many of the natives and their tribes, with the emergence of the ceremonial religious dances in the late 1880s, which have come to be known as the Ghost Dance. Getting its start in the Nevada territories, among the Paiute tribe, it spread among the Native American tribes within the central plains and beyond the Missouri River and Rockies. It was a messianic doctrinaire movement. Its founder was a 35-year-old known as Wovoka by his tribe and Jack Wilson by whites. His reputation as a medicine man had grown significantly among members of his tribe and the nearby surrounding tribes. He came down with a life-threatening fever. During his illness, and later convalescence, a brief eclipse of the sun infused his tribesmen and -women with a lot of excitement. In the throes of his life-threatening fever, Wovoka grew delirious and experienced an indelible image, which he strongly believed was a direct encounter with his ancestral god. In that state of delirium, Wovoka also felt that he had inhabited a purely spiritual realm. His ancestral god informed him that the final spiritual and secular redemption of himself and his people was very close at hand. He was told that their pillaged and stolen inheritance would be regained. His and fellow tribesmen's and -women's departed family members and friends would be united as one all-encompassing tribal family. Their ancestral simplicity, "good" and "happier" days, would be restored. The only matter at hand that they all needed to stay focused on and attend to was to perform in earnest the dance ceremonies and sing as one united people so that they could individually and collectively attain that moment of salvation they had been promised by their ancestral god and Wovoka. With regular performances of this dance, many hypnotic trances among the participants ensued. This became a very important coping mechanism, with all the misery that in time had come to envelope and surround them. In time, with its growing popularity, tribes such as the Sioux in the Dakotas, the Arapahos, and the Cheyenne embraced the song and dance ceremonies as their own. With the Sioux, the practicing of this ceremony helped to galvanize their tribesmen and -women through venting the anger they felt, regarding the murder of Sitting Bull on December 15, 1890. Then, exactly two weeks later, on December 29, the Wounded Knee massacre occurred. All-out conflict broke loose between the Sioux tribe and the federal forces. In time, the more virulent forms of the dance and song ceremonies receded in their overall impact on the tribes and the federal forces. But it persisted in more internal Native American social functions. The important fact regarding this unique native tribal event was that it was just another way by which the native populations, who had over the centuries become subjugated and marginalized by the larger European society, could vent their bitter and deep frustrations to the seemingly "immune" practices and powers imposed upon them by what they perceived to be a virulent and

destructive alien presence. This is also what had informed and driven the Pontiac Rebellion of 1763 and 1764, and immediately prior to the War of 1812, it emerged in Tecumseh and his brother Tenskwatawa's impact, influence, and eventual rebellion against federal and state forces by their local and surrounding tribesmen. That, according to diverse members of the indigenous populations, if anything, was the real "epidemic" they needed to be forever vigilant of and regularly combat—alien European powers and their proxy settlers and their combined persistent and avarice means of encroachment.

So, in the end it was not Eastern technology that allowed Europeans and their later American descendants to prosper and conquer the entirety of the American continents. The seeds of their dominance were, quite literally, sown even before the first permanent settlements were founded. With the first fish catches dried by English sailors on Newfoundland's shores, the biological exchange began. By the time the Pilgrims arrived at Plymouth, they found the perfect location for a settlement, seemingly abandoned by its native inhabitants. By the 18th and 19th centuries, that biological advantage had come to be internalized by American settlers moving west. They felt entitled to dominate the west, and promoters and social theorists were only too happy to lend their voice to support their cause. Despite the earlier "immunity" that Europeans might have felt toward all indigenous populations, not only in the Americas but throughout the world, they will not be immune from the judgment of future generations, native and European. The Atlantic slavery of the 16th through 19th centuries and the collectively accepted evils that it had perpetrated will be equated with the present and continued treatment of the descendants of the native populations that were conquered by earlier Europeans, and the fruits of the conquered populations' labors have and continue to be enjoyed by these earlier Europeans' descendants. Biological factors may have started Europeans down the road toward colonialism and domination of indigenous populations, but hubris continued it, and it is that hubris that largely forms the context for Native American relations today.

References and Further Reading

Berger, Peter L., and Thomas Luckmann. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Knopf Doubleday, 1967.

Bose, Sugata. *A Hundred Horizons: The Indian Ocean in the Age of Global Empire*. Cambridge, MA: Harvard University Press, 2009.

Crosby, Alfred W. *The Columbian Exchange: Biological and Cultural Consequences of 1492*. Westport, CT: Greenwood Press, 1972.

Crosby, Alfred W. *Ecological Imperialism: The Biological Expansion of Europe, 900–1900*. Studies in Environment and History Series. Cambridge: Cambridge University Press, 2003.

Crosby, Alfred W. "Virgin Soil Epidemics as a Factor in the Aboriginal Depopulation in America," *William and Mary Quarterly* 33 (April 1976): 289–299.

Curtin, Philip, and Edmund Burke III, eds. *Cross-Cultural Trade in World History*. Cambridge Studies in Comparative World History. Cambridge: Cambridge University Press, 1984.

Curtin, Philip, and Edmund Burke III, eds. *The Rise and Fall of the Plantation Complex: Essays in Atlantic History*. Cambridge Studies in Comparative World History Series. Cambridge: Cambridge University Press, 1998.

Emmer, Peter. "The Myth of Early Globalization: The Atlantic Economy, 1500–1800." *European Review* 11, 1 (2003): 37–47.

Geertz, Clifford. *The Interpretation of Cultures*. New York: Basic Books, 1997.

McNeill, John Robert. "Yellow Jack and Geopolitics: Environment, Epidemics, and the Struggles for Empire in the American Tropics, 1650–1825." *OAH Magazine of History* 18, 3 (April 2004): 9–13.

McNeill, William H. *Plagues and Peoples*. Garden City, NY: Anchor Press, 1976.

Klooster, Wim, and Alfred Padula, eds. *The Atlantic World: Essays on Slavery, Migration, and Imagination*. Upper Saddle River, NJ: Pearson/Prentice Hall, 2005.

Krige, John. *American Hegemony and the Postwar Reconstruction of Science in Europe*. Transformations: Studies in the History of Science and Technology. Cambridge, MA: MIT Press, 2006.

Mangan, J. A. *The Games Ethic and Imperialism: Aspects of the Diffusion of an Ideal*. Cambridge: Cambridge University Press, 1981.

Williams, William Appleman. *Empire as a Way of Life*. Oxford: Oxford University Press, 1980.

Williams, William Appleman. *The Tragedy of American Diplomacy*. New York: Norton, 1988.

Wright, Donald R. *The World and a Very Small Place in Africa: A History of Globalization in Niumi, the Gambia*. Armonk, NY: M.E. Sharpe, 2004.

6

Richard III was innocent of the charge of murder.

PRO	Charles Beem
CON	Jeffrey Mifflin

PRO

There is no positive or conclusive historical evidence of any kind to suggest that King Richard III of England (r. June 1483–August 1485) murdered his nephews, King Edward V (r. April–June 1483) and his younger brother, Richard, Duke of York, or that these princes necessarily died at all during Richard's reign. The young and tragic Edward V, whom Richard deposed after a reign of less than three months, has the most shadowy history of all the kings of England before or after him. In fact there are no contemporary witnesses to their alleged deaths. Instead all contemporary eyes and subsequent pens remained focused on the power struggle surrounding his throne and failed to notice how, when, or why he and his brother disappeared from sight following Richard's usurpation of the crown. Not surprisingly, as historian Alison Hanham once observed, "in general, it would be safe to say that our picture of English history in the later fifteenth century is deceptively clear because there are so few details to confuse the eye" (1975: 155). In the case of the probable deaths of the princes, however, there are no solid details whatsoever.

Nevertheless, accusations of Richard's guilt in their murder have resounded down the centuries, the centerpiece to his enduring popular reputation as England's most wicked king. However, the complete lack of contemporary witnesses describing how and why the princes disappeared renders accusations of Richard's guilt in their probable murders, all of which were made *after* his death at the Battle of Bosworth (August 22, 1485), mere smoking guns that, in more than one mock trial of recent times (1996 and 1997), have acquitted the last Yorkist king of direct complicity in the murder of the princes. Because Richard himself lost his crown to another usurper, Henry Tudor (subsequently King Henry VII), the succeeding Tudor Dynasty (1485–1603) had a vested interest in blackening Richard's character to bolster their own rather dubious legitimacy.

By the reign of Elizabeth I (1558–1603), this evolutionary process of historical character assassination had reached the proportions depicted in William Shakespeare's famed historical drama *The Tragedy of Richard III*, which regaled audiences with a grotesque, deformed, morally bankrupt prince out "to

prove a villain," who had already murdered scores of other individuals long before he got his hands on the unfortunate princes in the tower. As a host of modern historians have demonstrated, nothing could have been further from the truth. As the historical evidence is sifted and balanced, a clear historical case of *reasonable doubt* emerges concerning Richard's complicity in the possible deaths of the tragic and unfortunate princes in the tower.

The Historical Background

Before the death of his elder brother, Edward IV, the first Yorkist king (r. 1461–1470, 1471–1483), Richard, Duke of Gloucester, had been conspicuous in his unswerving loyalty to the interests of the York family. This was exemplary conduct in the context of the later 15th century in England and a highly prized commodity during the various events collectively referred to as the Wars of the Roses. These conflicts, a series of power struggles, street fights, and occasional pitched battles, occurred from 1455 to 1487 between the rival dynastic Houses of Lancaster and York and their erstwhile and eminently changeable allies among the nobility. This was a particularly violent and unstable period of English history for members of the aristocracy and gentry, as they competed for power, patronage, and influence within the treacherous confines of the royal court. Edward IV himself, the dazzling sun king of York, king at the age of 18 in 1461, had been particularly ruthless in the pursuit and maintenance of his royal estate, both deposing (twice, in 1461 and 1471) and probably ordering the murder of the decrepit Henry VI (r. 1422–1461, 1470–1471), the tragic and unfortunate final Lancastrian king, who supposedly expired in the Tower of London from "pure displeasure and melancholy."

During the first phase of Edward IV's reign (1461–1470), Gloucester emerged from adolescence as what biographer Paul Murray Kendall, Richard's most enthusiastic modern historical defender, once referred to as "the sturdy prop of his brother's throne" (1955: 514). This was in marked contrast to the career of the elder of Edward's surviving brothers, George, Duke of Clarence, who eventually came under the influence of Richard, "The Kingmaker," Earl of Warwick, whose initial influence over the king began to wane over the course of the 1460s. The quintessential "over-mighty subject" of the 15th century, Warwick convinced Clarence to marry his elder daughter Isabel in 1469, in flagrant violation of the king's firm refusal. By 1470 these two men entered into a plot that constituted the nadir of 15th-century English political history. In 1470, when Warwick, with a confused Clarence in tow, struck a bargain with Henry VI's queen, Margaret of Anjou, to toss Edward IV off his throne and replace him with the previously deposed Henry VI, Edward's prisoner in the Tower. Henry VI's "redemption" of 1470–1471, which placed a previously deposed, and utterly worthless, king back on his throne to play the role of royal puppet, was

the ultimate dirty deal of dynastic convenience between formerly implacable enemies who aptly represented the stunning decline of the respect for and sanctity of the royal office that occurred over the long expanse of Henry VI's disastrous adult reign. Through all the sudden changes of fortune that Edward IV endured, Gloucester remained perfectly loyal to his interests. Indeed before his bid to become king in 1483, Richard of Gloucester's career as an unwavering supporter of the House of York was without blemish and provided a startling and refreshing contrast to the morally bankrupt opportunism that characterized the political culture of his day.

Once back on his throne after crushing the Houses of Lancaster at the Battle of Teweksbury (May 4, 1471), Edward IV enjoyed relative stability as king for the remainder of his reign. Indeed he was a capable administrator and financier, at least when he applied himself to business, but Edward nevertheless sowed the seeds that ultimately served to destabilize the Yorkist succession. Three years after becoming king in 1464, Edward IV had impetuously married, well below his station, the Lancastrian widow Elizabeth Grey (née Wydeville), while his chief minister, the Earl of Warwick, had labored in ignorance negotiating a French marriage for the king. Despite two sons from her previous marriage, Thomas, later created Marquis of Dorset, and Richard Grey, Elizabeth Wydeville bore Edward four daughters and two sons. The eldest son, named Edward was created Prince of Wales soon after his birth and was sent to Wales to live under the tutelage of his uterine uncle, Anthony Wydeville, Lord Rivers.

Back in London, Edward presided over a court riddled by faction and intrigue. While Gloucester had endeavored to reconcile Clarence back to his brother in 1471, the Wydevilles had never forgiven Clarence for the role he had played in the death of the queen's father, the first Earl Rivers, in 1470, and worked unceasingly to destroy him. Following Clarence's execution (February 18, 1478), the Wydevilles, rapacious in their desires for aristocratic marriages, wealth, and power, competed for power and patronage at court with Edward's loyal household men and the rest of the nobility, whose ostensible leader was William, Lord Hastings. Though implacable enemies, Hastings and Dorset were both enthusiastic companions in Edward IV's lecherous and gluttonous extracurricular activities, indulging his passions and desires with striking regularity, which ultimately hastened the king's early demise.

Apart from these groupings in the final years of the reign stood Richard of Gloucester, a tower of moral rectitude, serving the king as royal viceroy in the north of England, a virtually autonomous jurisdiction that Edward IV bestowed on his younger brother on account of his leadership abilities and unwavering loyalty. Because of his complicated responsibilities in the last few years of Edward IV's reign, Gloucester had been an infrequent visitor to the king's court and stood aloof from the factional strike that he undoubtedly had no sympathy for whatsoever. However, there is no evidence to suggest that Gloucester was

on bad terms with the Wydevilles, members of Edward's royal household, or the rest of the Yorkist nobility.

When Edward IV's health began to fail rapidly at the beginning of April 1483, he fully realized the dangerous lack of consensus among the various factions present at his court, which his deathbed exhortations were unable to reconcile. Edward's imminent death meant the onset of a royal minority, that is, the reign of an underage king, which required some form of regency. The most successful of medieval minorities were those of Henry III (1216–1227), which enjoyed the singular leadership of the legendary William Marshal, the Earl of Pembroke, the loyal servant of the first four Plantagenet kings, and Henry VI (1422–1437), created king at the age of nine months, whose minority succeeded because the lords temporal and spiritual worked together consensually and corporately to govern the realm during the young king's protracted minority. While the 1422 precedent pointed toward a conciliar form of regency, in 1483, as in 1216, only one man, Gloucester, could possibly have imposed consensus on the factional strife that was endemic at the Yorkist court. Thus the king probably put much faith into the last-minute codicils added to his will that appointed Gloucester as protector of the king and kingdom during the inevitable minority of his heir. When Edward IV died on April 9, 1483, however, the new king, 12-year-old Edward V, and his protector and uncle, Richard of Gloucester, were both equidistant from London, in Wales and Yorkshire, respectively.

The Prosecution's Case

In the capital, Queen Elizabeth and Hastings squared off for control of the minority regime, as the queen wrote to her brother and son in Wales, urging them to reach the capital as quickly as possible. The queen did not inform Gloucester of her actions, in the belief that if the new king could be crowned before Gloucester reached London, this would nullify his protectorship, in accordance with the 1429 precedent, when the protectorship of Humphrey, Duke of Gloucester, was terminated following seven-year-old Henry VI's English coronation. An expedited coronation sought to circumvent Gloucester's appointment as protector and place de facto power in the hands of the queen and her adherents. Despite opposition from Edward's household men, the queen was able to dominate the initial meetings of a royal council that had come together spontaneously upon Edward IV's death. As Michael Hicks (1979) has argued, the reign opened with a Wydeville coup d'état, as the queen prevailed on the council to endorse her suggestion for a hasty coronation, scheduled for May 4. Gloucester, however, had been informed of the queen's machinations and wrote to Earl Rivers in Wales, asking if their parties could meet, so the protector could personally escort the new king into London, a request that Rivers, not suspecting any subterfuge, could hardly have declined.

However, when the rendezvous occurred in Northampton (April 29, 1483), Gloucester only found Rivers; the king was 14 miles down the road closer to London in Stoney Stratford. Joined by Henry Stafford, Duke of Buckingham, a kinsmen of the Lancastrians who had been forced to marry one of the queen's sisters, Gloucester enjoyed a convivial evening with Rivers and the next morning promptly arrested him, the king's half-brother Richard, and Thomas Vaughan and shipped them off to the north of England for safekeeping. Gloucester and Buckingham then proceeded to Stoney Stratford, where Gloucester took physical possession of the king, the first positive step toward Edward V's deposition. Richard's historical prosecutors have often claimed that at this moment Richard was already scheming to obtain the crown.

Despite the naked show of force, Gloucester still enjoyed considerable support among the ranks of Yorkist political society. Before arriving in the capital, Gloucester had written sober letters defending his conduct while accusing the queen and her kindred of conspiring against him. Gloucester and the king entered London on May 4, 1483, in an impressive ceremony, while Queen Elizabeth took her daughters, younger son, and all her movables with her into the sanctuary of Westminster Abbey, after failing to attract support for an armed force to challenge Gloucester's possession of the king. The majority of the dead king's councilors, including Hastings, initially approved of Gloucester's actions and quickly affirmed his appointment as protector, giving him, according to the *Croyland Chronicle*, "power to order and forbid in any matter, just like another king" (cited in Pronay 1986). Meanwhile the young king himself, treated with all due deference by Gloucester and the rest of the nobility, was lodged first in the London house of the bishop of London and later in the royal apartments of the Tower of London, as he awaited his coronation, set for June 22.

Once installed as protector, Gloucester's right-hand man turned out to be Buckingham and not Hastings, the late king's closest friend and confidante, who apparently expected a significant quid pro quo from Gloucester for his earlier support. Hastings was
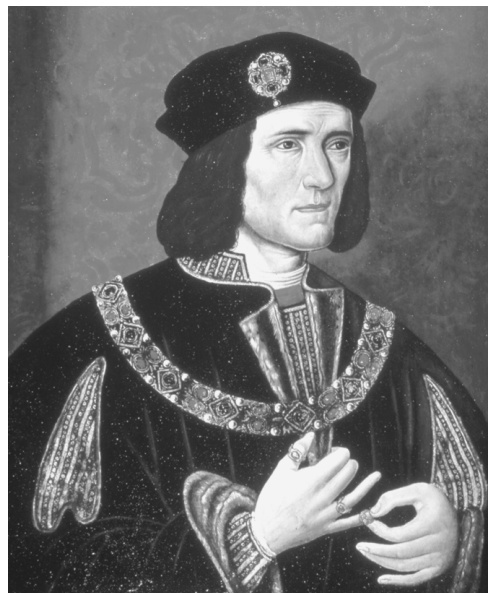
By deposing and then perhaps murdering his nephew Edward V and his younger brother, Richard III revived the Wars of the Roses, thereby destroying himself and his dynasty and making possible the rule of the House of Tudor under King Henry VII. (Corel)

apparently resentful, remaining outwardly cordial while maintaining personal access to the king in the tower. If not before, by Friday, June 13, Gloucester had irrevocably decided to usurp the crown. At a meeting of the council in the tower on that morning, Gloucester accused Hastings; Thomas Rotherham, archbishop of York; and John Morton, bishop of Ely, of conspiring with the Wydevilles to gain possession of the king. Hastings was hustled out to the Tower Green and summarily beheaded, while Rotherham and Morton were taken into custody. By now an overawed council agreed with the protector's demands that the king should not be crowned without the presence of his younger brother, the Duke of York, still in the sanctuary of Westminster Abbey with his mother and sisters. On June 16, Thomas Bourchier, archbishop of Canterbury, removed the young prince from the abbey and reunited him with his brother in the tower. At the same time, Gloucester secured custody of Clarence's six-year-old son, the Earl of Warwick, the only other direct male descendant of the House of York.

With all the male scions of the House of York in his possession, Gloucester set his usurpation into motion. On June 17, parliament was canceled as plans for the coronation were abandoned. The pretext that Gloucester chose as "colour for this act of usurpation" was the "discovery" that Edward IV's children were illegitimate, because, according to Robert Stillington, bishop of Bath and Wells, a longtime adherent of Clarence and foe of the Wydevilles, Edward had entered into a binding precontract with lady Eleanor Butler before his marriage to the queen. On Sunday, June 22, Gloucester and a host of other notables came to Saint Paul's Cross to hear friar Ralph Shaa preach that "bastard slips shall not take root." Three days later, Gloucester mounted the throne as Richard III. By this time, Earl Rivers and the king's half brother, Richard Grey, had also been summarily executed.

While Richard had indeed been ruthless in his dogged pursuit of the crown in the face of allegedly relentless factional hostility, what exactly happened to

---

The Tower of London

The Tower of London is actually a multibuilding complex north of the Thames River and just east of London. Originally built as a fortress in the 11th century, the tower is protected by a moat and two defensive walls. The tower has also long been the home to wild ravens, which in modern times are cared for by the royal ravenmaster, the bulk of the raven population having died of shock during the World War II bombing of London.

Over the centuries, the tower has been used as a prison and a place of torture for dissidents. In addition to the princes, other famous prisoners have included Queen Elizabeth I, Guy Fawkes, Sir Walter Raleigh, Thomas More, and Anne and Jane Boleyn. The ghost of Anne Boleyn is said to wander the grounds of the tower, carrying her severed head.

the princes nonetheless remains English history's greatest unsolved mystery. Shortly after his uncle seized the throne, the deposed Edward V and his brother completely disappeared from view and contemporary comment. Dominic Mancini, an Italian cleric visiting London during much of 1483, reported that the princes were still seen shooting butts on the Tower Green at the end of June and related how the ex-king's physician, Augustine, reported that "the young king, like a victim prepared for sacrifice, sought remission of sins by daily confession and penance, because he believed that death was facing him" (Mancini 1936). Mancini's observations, however, are the last recorded contemporary sightings of the princes, who henceforth disappeared completely from the historical record. At this point, the prosecution has failed to produce a *corpus delecti*, the fact of a crime actually being committed.

The Defense

Following Richard's coronation and subsequent reign, the princes are simply not mentioned in the historical record. Only after Richard had died at Bosworth Field did any contemporaries speculate on the likely fate of the princes. But apparently by the fall of 1483, it was widely *rumored* that the princes were dead, which served as the catalyst for Buckingham's rebellion in October, fought, according to the *Croyland Chronicle*, to place the exiled Henry Tudor, Earl of Richmond, a descendant of the Lancastrian royal house, on the throne, because Tudor vowed to marry Edward IV's eldest daughter, Elizabeth of York, and unite the two rival dynasties. It is entirely possible that Buckingham, who had enjoyed great favor from Richard before his rebellion, a situation that perhaps had encouraged Hastings's alleged treason, arranged for the princes to be murdered to smooth Tudor's, or even perhaps his own, path to the throne. The continuation of the *Chronicle* describing Richard's reign, written early in the succeeding reign of Henry VII, mentioned that Buckingham repented of his support for Richard III and left London at the same time that rumors began circulating concerning the princes' fate. However, Richard swiftly crushed Buckingham's rebellion and beheaded its ringleader without mercy, who took whatever secrets he had concerning the fate of the princes with him to his grave.

If dynastic security was the primary motivation for the actual murder of the princes (which, remember, has never actually been proven!), then Buckingham and Tudor had just as much of a vested interest in seeing the princes succumb to unnatural deaths as did Richard III. Tudor-era historian Polydore Vergil, in his *Anglica Historia*, described Buckingham's dynastic pretensions in a particularly lively section of his narrative, as Richard III accused Buckingham: "What now, duke Henry, will yow chalenge unto you that right of Henry the Fourth wherby he wyckedly usurpid the crowne, and so make open for yourself the way therunto?" (Vergil 1555). What all this amounts to is a smattering of circumstantial

evidence against Richard, enough to indict, but far short of an actual conviction, as both Buckingham and Tudor also look like probable suspects.

As Buckingham and Tudor are just as likely assailants as King Richard, the dowager queen, Elizabeth Wydeville, provides highly contradictory and unreliable testimony and behavior that continues to befuddle scenarios envisioning Richard's guilt in the murder of the princes. Were she alive, even if she had knowledge of potentially damning testimony, the prosecution would probably have declined to call her to the witness stand, because her testimony (and character) would have been destroyed under cross-examination. Before Buckingham's rebellion, Elizabeth had agreed to a dynastic deal hatched by Margaret Beaufort, Tudor's mother, for the marriage of Henry Tudor and Elizabeth of York, which suggests that she was aware or suspected that her sons were dead. If, in fact, she knew her sons were dead, Elizabeth remained focused on securing her surviving children's dynastic rights, despite the vividly described grieving mother scene in the narrative of Thomas More's (2005) *The History of King Richard III*.

Following the failure of Buckingham's rebellion, the queen's emotionless, realpolitik attitude was fully revealed on March 1, 1484, when she and her daughters left sanctuary and were reconciled to King Richard and his royal court. There are several plausible explanations for her rapprochement with the man who had executed her brother and son and who was widely rumored to have instigated the deaths her two younger sons as well. One is that Elizabeth knew the princes were still alive, or had died natural deaths. The other is that she was simply tired of remaining in sanctuary and accepted the reality of the present administration. Elizabeth lived until 1492, but no viable contemporary or later source mentions any of her views regarding the fate of her two younger sons. Elizabeth did not apparently enjoy a cordial relation with her son-in-law, Henry VII, who shut her up in the Abbey of Bermondsey in 1487, after she had inexplicably lent her support to the Yorkist pretender, Lambert Simnel, which defies any rational interpretation of her possible motivations. Like the Duke of Buckingham, Elizabeth Wydeville went to her grave perhaps with some knowledge of the true story of the fate of her sons.

As there are no contemporary accounts of what exactly happened to the princes, all subsequent accounts of Richard's supposed guilt in the fate of the princes were manufactured after his death, during the reigns of the first two Tudor kings, Henry VII and VIII, long after the supposed "fact" of the prince's demise sometime in the later half of 1483. While centuries of historians have uncritically accepted Thomas More's interpretation of a physically and morally corrupt, infanticidal tyrant, 20th-century historians have successfully reopened the case and cast grave doubts on nearly all of More's assertions concerning Richard's character and crimes, including the murder of the princes he so richly described.

Richard's modern defense reached the larger popular culture with the publication of Josephine Tey's 1951 novel, *The Daughter of Time*, which, among other

things, exposes the weaknesses and biases of More's *History*, a narrative allegedly based in part on the personal recollections of John Morton, formerly bishop of Ely. Morton escaped confinement during Buckingham's rebellion late in 1483, having been incarcerated following the June 13 coup, which resulted in the summary beheading of Hastings, and made his way to Henry Tudor in Brittany. Described by Dominic Mancini as a man "of great resource and daring, for he had been trained in party intrigue since King Henry's [VI] time" (1936), Morton emerged in Henry VII's reign as lord chancellor, cardinal, and perhaps the principal architect of Richard III's historical character assassination. More's *History* has also been explained as a rhetorical, humanistic literary exercise, rather than a precise history, perhaps offering to the youthful Henry VIII an example of how a prince ought *not* to behave. In the words of Jeremy Potter, More's work was "a literary exercise in the dramatic presentation of villainy" (1991) that bore little relationship to the historical King Richard.

More's book, however, is the first to describe Sir James Tyrell's alleged murder of the princes, a story reproduced and embellished in every subsequent work of 16th-century history, through the chroniclers Edward Hall, Ralph Holinshed, and finally, the bard of Avon himself, William Shakespeare. What More did not reveal, however, was the contradictory and shady relationship that existed between James Tyrell and Henry VII.

Tyrell, in fact, enjoyed great favor under Henry VII, recovering the offices he held under Richard III and enjoying frequent employment in the king's service. In June and July 1486, Henry VII issued to Tyrell two pardons for unspecified offences without explanation. Why would Henry VII refrain from publicly acknowledging the murder of the princes if he knew Tyrell was in fact the murderer? Unless, of course, it was Tudor, not Richard, who had arranged for the princes to be done away with. Only later in the reign, in 1502, when Tyrell had provided aid and comfort to the renegade Yorkist prince Edmund, Earl of Suffolk, in his capacity as lieutenant of Guisnes, was he convicted and executed for treason. How and why Henry VII allowed Tyrell's supposed guilt in the princes' murder to be known, if he in fact did anything of the sort, long after the fact of the deed, remains a mystery, as does how More himself got hold of the story. Polydore Vergil failed to relate Tyrell's supposed confession, as did Bernard Andre, tutor to Henry VII's sons and the author of his biography. Writing in the early 17th century, Francis Bacon let it be known that Henry VII allowed knowledge of Tyrell's guilt to be circulated. Ultimately the full historical disclosure of the career of Tyrell, and of his highly suspicious relationship with Henry VII, only serves to cast further doubt on Richard as suspect number one.

However, before we proceed to our closing arguments, we must admit one final piece of evidence, which was the discovery of the bones of two preteen children found under a staircase in the Tower of London in 1674. The location roughly fits with Thomas More's account of the disposal of the princes' bodies.

King Charles II had the remains interred in an urn in Westminster Abbey, where they have remained ever since, which has further contributed to Richard's enduring reputation as their murderer. In 1933, however, the remains were examined by scientists armed with the rudimentary forensic science of their day, who rendered an inconclusive verdict regarding their identity or their sex. Since that time, subsequent British monarchs have forbid further scientific analysis of the remains. It is unlikely that the monarchy will ever allow their examination again, despite the state of forensic science today, which could easily identify their DNA.

Conclusion

Ultimately what all the admissible evidence adds up to is an ironclad case of *reasonable doubt*. No contemporary account points the finger directly at Richard for the murder of the princes; only subsequent accounts labor to convince us of his guilt. However, considering his bloody path to the throne, Richard remains the most likely circumstantial candidate. Like a medieval O. J. Simpson, Richard's repeated acquittals in courts of law do not erase the strong presumption of unproven guilt in the larger court of public and historical opinion. As historian Charles Ross once pointed out, even Paul Murray Kendall, Richard's most vigorous modern defender, admits "the most powerful indictment of Richard is the plain and massive fact that the princes disappeared from view after he assumed the throne, and were never reported to have been seen alive. This fact . . . weighs heavily against the indications of his innocence" (1981).

If he was guilty, Richard did have a large body of powerful and bloody precedents to draw on: Edward II's probable murder at the hands of his queen, Isabella, and her paramour, Roger De Mortimer, in 1327; Thomas of Woodstock's probable murder on the orders of Richard II in 1397; and Richard II's own untimely demise on the probable orders of his supplanter, Henry IV, the first Lancastrian king, in 1400.

The precedents closer in time to 1483 were just as ruthless and far more numerous. Henry VI's ferocious queen, Margaret of Anjou, beheaded every Yorkist lord she could get her hands on and placed a paper crown on the severed head of Richard, Duke of York, following the Lancastrian victory at 1460 Battle of Wakefield. Edward IV dispatched the decrepit old Henry VI 11 years later following the 1471 Battle of Teweksbury after Henry's son, 17-year-old Edward of Lancaster, had perished during that battle. Indeed, if Edward had survived Teweksbury, Edward IV could hardly have allowed the youth to live without putting his recently regained crown at continued dynastic risk. Perceiving the same threat, the Wydevilles succeeded in destroying George, Duke of Clarence. This was the political culture in which all of our suspects were constrained to participate in. In 1499, 14 years after his accession, Richard's supplanter, Henry

VII, beheaded the hapless Edward, Earl of Warwick, son of George, Duke of Clarence, and the last direct male descendant of the House of York, during the negotiations for the betrothal of his eldest son and heir, Arthur, prince of Wales, and Catherine of Aragon, youngest daughter of Ferdinand and Isabella of Spain.

Nevertheless, despite the circumstantial evidence that weighs heavily against him, there is no contemporary account of the fate of the princes. Dominic Mancini summed it up best in his account of the usurpation and subsequent disappearance of Edward V, as he wrote, "I have seen many men burst forth into tears and lamentations when mention was made of him after his removal from men's sights; and already there was a suspicion that he had been done away with, and by what manner of death, so far I have not at all discovered" (1936).

References and Further Reading

Bruce, John, ed. *History of the Arrivall of Edward IV, in England and the Finall Recouerye of His Kingdomes from Henry VI A.D. M.CCCC.LXXI*. London: Camden Society, 1838.

Hanham, Alison. *Richard III and His Early Historians*. Oxford, UK: Clarendon Press, 1975.

Hicks, Michael. "The Changing Role of the Wydeville's in Yorkist Politics to 1483." In *Patronage, Pedigree, and Power*. Edited by Charles Ross. New York: Alan Sutton, Rowman and Littlefield, 1979.

Kendall, Paul Murray. *Richard the Third*. London: George Allen and Unwin, 1955.

Mancini, Dominic. *The Usurpation of Richard III*. Edited by C. A. J. Armstrong. Oxford, UK: Clarendon Press, 1936.

More, Thomas. *The History of King Richard III*. Edited by George M. Logan, Bloomington: Indiana University Press, 2005.

Potter, Jeremy. *Good King Richard? An Account of Richard III and His Reputation*. London: Constable, 1983.

Potter, Jeremy. "Richard III's Historians: Adverse and Favorable Views." Richard III Society (1991). Available online at http://www.r3.org/bookcase/misc/potter.html (accessed August 10, 2010).

Pronay, Nicholas, and John Cox, eds. *Crowland Chronicle Continuations, 1459–1486*. London: Alan Sutton, 1986.

Ross, Charles. *Richard III*. Berkeley: University of California Press, 1981.

Shakespeare, William. *The Tragedy of King Richard III*. Edited by John Jowett. New York: Oxford University Press, 2001.

Tey, Josephine. *The Daughter of Time*. New York: Macmillan, 1951.

Vergil, Polydore. *Anglica Historia*, London: 1555.

CON

A prolonged dynastic struggle for kingship ripped through England in the 15th century. The Wars of the Roses (1455–1485) pitted the House of Lancaster (the emblem of which was a red rose) against the House of York (symbolized by a white rose). The throne passed from house to house, and deceptions, betrayals, judicial executions, and murders were among the tactics employed to obtain and secure power. Among the Yorkists was Richard, Duke of Gloucester, younger brother of King Edward IV. Richard seized power soon after his brother's death in 1483 despite the fact that Edward IV had two living sons (Edward and Richard) who were in line to inherit the throne. Richard of Gloucester became Richard III. The Wars of the Roses ended in 1485 when Henry Tudor's forces defeated those of Richard III at the Battle of Bosworth. Henry became King Henry VII, founder of the Tudor Dynasty. Richard III's battered remains were ignominiously trussed over a horse's back and carried off to an obscure burial without honors.

Richard III's reign lasted only 25 months, but his notoriety was such that his name is, even today, among the most recognized of any English monarch. He may be the most controversial ruler England has ever had, exceeding in that regard even King John. It is true, as his rehabilitators point out, that Richard III has been unjustly maligned. It is highly unlikely that he was born after two years in his mother's womb with a full set of teeth, or that he had a predilection in infancy for devouring frogs. The death of saintly, feeble Henry VI in the Tower of London cannot justifiably be ascribed to Richard, even though a 1910 examination of Henry's skull found it to be "much broken," suggesting that his unhappy end was precipitated by something other than the "pure displeasure and melancholy" mentioned in Yorkist propaganda. Nor did Richard III stab Prince Edward of Lancaster to death. Edward (Henry VI's son) died fighting at the battle of Tewkesbury. It is no longer the current view that Richard III drowned his brother, George, Duke of Clarence, in a butt of Malmsey, nor is it likely that he murdered him by any other means.

But acquitting Richard III of the aforementioned three deaths still leaves him accountable for the untimely demise of William, Lord Hastings; Richard Grey; Sir Richard Haute; Anthony Woodville, Earl Rivers; Sir Thomas Vaughan; Sir Thomas Saint Leger; William Colyngbourne; and Henry Stafford, Duke of Buckingham, all of whom, according to Charles Wood (1978) in "Who Killed the Little Princes in the Tower," lost their heads because Richard feared conspiracies against his power and person. As we shall see, the death of the little princes fits neatly into the pattern of elimination used by Richard III against those thought likely to pose a threat.

One of the key events of Richard III's reign was the sequestration in the summer of 1483 of his nephews, Edward V and Richard, Duke of York, in the Tower of London. The little princes, aged 9 and 12 when they were shut away, disappeared mysteriously while under their uncle's protection. Rumors flew that

View of England's Tower of London. Built in the 11th century, the Tower of London has served as a royal residence and a prison, as well as a site for beheadings and hangings. (Philip Lange/Dreamstime.com)

Richard had arranged their deaths to consolidate his power, and no contemporary writer expressed a different opinion. Historians who wrote during the reign of Henry VII were even more emphatic in their accusations.

"What is so special," asks Jeremy Potter, former chairman of the Richard III Society, "about Richard III's alleged villainy when compared with that of, say, Henry IV, Edward IV, and Henry VII—to name but three usurping regicides—or numerous other bloodstained 15th-century toughies?" (cited in Wood 1978). The answer, of course, is that these were the forlorn children of his brother who were, supposedly, under his protection.

Modern apologists for Richard, in attempts to rehabilitate his reputation, have engendered a debate with historians who subscribe to the traditional view that he caused the murder of his nephews. They urge that Richard was innocent. He did not, they assert, have a motive for the murders and could not have been guilty of ordering such a heinous act against his beloved brother's sons. They point to bits and pieces of historical evidence and conjectures derived from them in their less-than-convincing exculpatory attempts. The usual apologist view is that Henry VII ordered the murders at some point early in his reign after Richard III had been killed at Bosworth. Even if, they continue, the little princes died during the reign of Richard III, the reprehensible deed was accomplished without his knowledge and contrary to his wishes.

This section will take the traditional point of view that Richard III, aware that his claims on the throne were tenuous and that his power could be easily undermined by an uprising consolidated around a figure more popular than himself, was guilty of arranging the murder of the little princes in the Tower of London.

Political and Historical Background to the Murders

Richard of Gloucester, as well as his most trusted advisers, hailed from the north of England. In southern England, where London and the royal court were located, people viewed the north with singular distrust. Northerners were considered wild, bellicose, licentious, and unfamiliar with the culture of the capital. Richard came south after the death of Edward IV (April 9, 1483) to keep an eye on his brother's minor children, especially Edward, the oldest of the two princes, who reigned briefly without being crowned as King Edward V.

Edward IV's queen was Elizabeth Wydeville, mother of the little princes and their sisters. Soon after Edward IV's demise, Richard dismissed the friends and advisers surrounding Edward V (on April 30) and had his own men escort the young king to London, where he was housed in apartments at the Tower of London. Elizabeth, highly alarmed, took refuge with her remaining children at Westminster Abbey. Richard officially became Edward V's protector on May 10. He persuaded Elizabeth to allow Edward's younger brother, Richard, Duke of York, to join him at London Tower on the pretense that Edward was lonesome and needed a playmate. He arranged his own coronation (July 6) as King Richard III instead of crowning Edward V, the rightful heir. Richard's pretense for seizing the throne was that the children of Edward IV were illegitimate and therefore legally unable to succeed to the kingship. This ruse was based on an allegation that Edward IV had entered into a "troth plight," a precontract for marriage, with Lady Eleanor Butler before his marriage to Elizabeth Wydeville, rendering invalid the marriage from which the little princes were born.

By midsummer, Richard's spies had been alerted to some sort of plot, the exact nature of which was not known. A faction favoring the restoration of Edward V reportedly wanted to spirit Edward's sisters to a safe haven overseas because of concerns that some malefactor might eliminate the male heirs. (If the male issue of Edward IV died, whatever rights they possessed to the crown would pass to their oldest sister, Elizabeth of York.) Richard responded by encircling Westminster Abbey and environs, where Elizabeth Wydeville and her daughters were in sanctuary, with armed men.

In late July, Richard III embarked on a royal "progress" to portions of the kingdom in an effort to win friends and solidify his support. The country acquiesced in Richard's usurpation, but acquiescence was not by any means in 15th-century England a guarantee of stability. Those who understood the law knew well that the petition deposing Edward V was not legally valid. If Edward IV's

marriage to Elizabeth Wydeville was null and void on the grounds of a previous "troth plight," it needed to be proved by means of a thorough investigation conducted by the church. Such disputes relating to marriage contracts were not unusual in the 15th century, and Richard and his advisers probably foresaw that canon lawyers, if given the chance, would decide in favor of Edward IV's children. He must have realized that many citizens, especially in the south, still considered Edward V the rightful king, even after the dissemination of convoluted stories about illegitimacy. Richard was apprehensive even in his better moments and must have realized that, while the princes were alive, his throne was insecure and his person was at risk.

Richard III's decision to permanently eliminate the threat posed by his nephews may have occurred even before he usurped the throne. It certainly lodged firmly in his mind once the rumblings of rebellion stirred shortly after his July 6, 1483, coronation.

Contemporary and Near Contemporary Accounts Relating to the Fate of the Princes

It isn't necessary to rely on accounts composed in the Tudor era, which have often been suspected of bias against Richard III, to sort out what happened in the ensuing months. Historians who want to cut through the myths surrounding Richard should look first at contemporary and near-contemporary records, histories, and other available documents. A number of documents, including reports and chronicles written before Henry VII's ascendancy in 1485 (or soon thereafter) have survived. The observations and allegations they contain can be studied and evaluated. Who were these authors and why should we lend credence to their accounts?

An Italian cleric and humanist writer named Dominic Mancini penned the most substantial contemporary narrative of events surrounding the disappearance of the little princes. He came to England in 1482 as a member of a diplomatic mission and wrote in Latin under the name Dominicus Mancinus. Living in London during the turbulent months following Edward IV's death, he had ample opportunities to observe, listen, and learn. His patron, Angelo Cato, archbishop of Vienne, recalled him around the time of Richard III's coronation. Mancini left London on July 6, 1483. Cato asked Mancini to record his impressions of the troubled situation in England. The resulting manuscript, an indispensable source for the events of April to July 1483, was completed in December, while his memories were still fresh. He recorded that Richard was adept at dissimulation, extremely ambitious, and lusted uncontrollably for power. He was, Mancini wrote, ruthless in his disposal of men, like Hastings, who stood in his way. His ruthlessness extended to the usurpation of his own nephew's crown.

Mancini noted that as soon as Hastings was beheaded on June 20, servants formerly appointed to take care of the little princes were "debarred access." The

boys, no longer seen playing in the tower gardens, "were withdrawn into the inner apartments of the Tower proper and day by day began to be seen more rarely behind the bars and windows, till at length they ceased to appear altogether. . . . The physician Argentine, the last of his attendants whose services the king enjoyed, reported that the young king, like a victim prepared for sacrifice, sought remission of his sins by daily confession and penance, because he believed that death was facing him" (Mancinus 1969). Mancini reported that even in July there were widespread suspicions that Edward V had been murdered.

A collection of signet letters and memorandums about political participants and the sequence of events during the protectorate and reign of Richard have come down to us and are preserved in the British Library as "Harleian Manuscript 433." A document therein dated July 14, 1483, authorized payment of wages (the last such payment) to 14 men for their services to Edward IV and to "Edward Bastard, late called King Edward V."

Richard III's career as a whole, including his short reign, is described in a contemporary narrative written at the Benedictine abbey of Crowland in Lincolnshire by an unidentified hand. Although the identity of the author has never been positively confirmed, it is obvious that he was intelligent and well informed about contemporary events. He is sometimes thought to have been John Russell, bishop of Lincoln, a doctor of canon law. Russell was Edward IV's keeper of the privy seal in the 1470s and served as Richard III's chancellor from 1483 to 1485. He was present at Crowland in April 1486. Entries in the *Crowland Chronicle* dealing with Richard III, known as the "second continuation," were composed over 10 days at the end of April that year. The author of the *Crowland Chronicle Continuation* wrote that in 1483 rumors were circulating that the little princes had come to a violent end and that their sisters might be taken overseas in disguise to save them from a similar fate. A later addition specifically accused Richard III of killing "his brother's progeny" (Pronay and Cox 1986).

Other contemporary and near-contemporary evidence is unanimous in corroborating that the rumors circulating in the summer and fall of 1483 about foul play at the Tower of London were essentially true. The *Great Chronicle of London* (1938), a manuscript at London's Guildhall Library, notes that in 1484, after Easter, "much whysperyng was among the people that the kyng had put the childyr of King Edward to deth." Philippe de Commynes, writing in 1498 at the behest of Angelo Cato, recorded his memories about the court of Louis XI of France. Louis XI, who supported a wide and reputedly efficient network of spies, was convinced by mid-August 1483 that Richard III had arranged for and accomplished the murder of his nephews. On January 15, 1484, Guillaume de Rochefort, the chancellor of France, warned the Estates General about the dangers of having a child king (Charles VIII was then only 14). He asked the Estates to remember the fate of Edward IV's children, who were "put to death with impunity, and the royal crown transferred to their murderer." (Dockray 1997)

Thomas More's Sifting and Weighing of Reports about the Princes' Deaths

Richard's rehabilitators have sometimes disparaged Thomas More's biography of Richard III as mere Tudor propaganda. But More (who later was made a saint) was a professional lawyer with considerable experience in investigating crimes, weighing evidence, and evaluating probabilities. Although he was just a child in 1483, he knew many people familiar with the events of Richard III's reign. Among those in More's political circles were men who had inside information about late-15th-century political events. He was aware of conflicting versions and went to some lengths to sort out the facts and reconstruct what probably happened in his *The History of King Richard III*. "I shall rehearse you the dolorous end," he writes, "of those babes, not after every way that I have heard, but after that way . . . heard by such men and by such means as methinketh [it would be difficult for it not to] be true" (More 2005: 97).

More started working on his manuscript around 1513 but left it unfinished and unpublished. There is no reason to suppose that More was influenced by Tudor pressures to malign Richard III during its composition. His narrative does contain some demonstrable errors, and many of the speeches he includes are imaginative reconstructions. In this practice he followed the example of classical authors such as Sallust and Tacitus. Like other Renaissance scholars, More looked to classical models. He was virulently opposed to tyranny, and his depiction of Richard III sometimes portrays the king as a dramatic stereotype of oppressive evil. His comparison, for example, of Edward IV as a good ruler with Richard III as a bad ruler is redolent of passages in Tacitus in which the Emperor Augustus is compared to his unworthy successor, Tiberius. Such dramatic flourishes should not get in the way of a balanced assessment and

---

Thomas More

Thomas More (1478–1535) was an English politician and author whose biography of Richard III was the principal source for Shakespeare's play of the same name. He's best known for his book *Utopia*, a word he coined for an ideal society; though the book is meant to be a satire, it's not always successful, and the society he described has been as much an influence as his mockery of it. Even his Utopian alphabet, a simplified script reminiscent in intent of Mark Twain's later satire of spelling reforms, informed the development of shorthand.

A pious Catholic, More opposed Henry VIII's assumption of control over the Church of England and had campaigned against the Protestantism of Martin Luther and others. When he refused to swear allegiance to the newly crowned Anne Boleyn by means of an antipapal oath, he was convicted of high treason and executed.

appreciation of the information that More collected, upon which the history is based. Historians should at least give serious consideration to More's basic characterizations and narrative of events, despite the presence of flaws. His account, in its broader outlines, is probably very close to the truth.

The reported facts, as sifted and arranged by More, are as follows. Richard III, following his coronation, set off on a "progress," touring parts of the realm accompanied by an entourage, for the purpose of showing himself to the people in hopes of earning their affection and gaining their loyalty. During his stay at Warwick, which he reached on August 8, Richard gave orders that the princes needed to be eliminated. He first sent a messenger to Robert Brackenbury, constable of the tower, directing him to put the children to death. The constable, although one of Richard's handpicked men, demurred, being too cautious or too overwhelmed by the enormity of the act. "Ah, who shall a man trust?" Richard exclaimed. One of his spies recommended James Tyrell, a knight who had performed dangerous and secret work for Richard on a previous occasion. Richard gave Tyrell a letter addressed to Brackenbury, ordering the constable to surrender the tower's keys for one night. Tyrell decided that the princes could be murdered most easily and discretely in their beds. He enlisted the help of Miles Forest, an experienced murderer, and his own horse keeper, John Dighton. "[T]his Miles Forest and John Dighton about midnight (the innocent children lying in their beds) came into the chamber and suddenly lapped them up among the clothes— so bewrapped them and entangled them, keeping down by force the featherbed and pillows hard unto their mouths . . . that smothered and stifled, their breth failing, they gave up to God their innocent souls . . ., leaving to their tormentors their bodies dead in the bed." The henchmen then fetched Tyrell to prove that the princes were dead. He "caused those murderers to bury them at the stairfoot, meetly deep in the ground under a great heap of stones" (More 2005: 100).

Forensic Corroboration

In 1674 workmen digging during renovations at the tower unearthed the skeletal remains of two children buried in a box under or near a staircase in the White Tower (a fortified keep within the larger enclosure). The bones were assumed to belong to the little princes who had disappeared without a trace 191 years before. King Charles II ordered the bones removed from the tower and had them ceremoniously deposited in Henry VII's chapel at Westminster Abbey in a sealed urn designed by the architect Christopher Wren.

In 1933 Lawrence Tanner, the archivist of Westminster Abbey, and Professor William Wright secured permission to unseal the urn and examine its contents in the presence of a dental surgeon, George Northcroft, who specialized in the treatment of children's teeth. They discovered that many bones were missing from the human skeletons and that the bones of various animals were mixed

in. A fairly complete human skull and another fragmentary human skull were in the urn, confirming the presence of two humans.

Teeth and bones were thoroughly examined. According to Lawrence Tanner and William Wright in "Recent Investigations," the researchers determined that the smaller remains were those of a child between 9 and 11 years old and that the remains showing more advanced growth belonged to a child between the ages of 12 and 13. A noteworthy characteristic of the remains of the older child was that an extensive stain, blood red above and brown below, fluid in origin, had discolored the area from the orbits to the angles of the lower jaw. It was, the researchers, concluded, a bloodstain. The parts of the facial skeleton had been separated, lending credence to the traditional account of the deaths, that the boys had been smothered under a feather bed and pillows and that the murderers had pressed down hard on their mouths. "The evidence that the bones in the urn are those of the princes is . . . as conclusive as could be desired," the researchers reported, "[and] their ages were such that their death occurred during the reign of their usurping uncle, Richard III. . . . We can say with confidence that by no possibility could either, or both, have been still alive on the 22nd August 1485, the date of Henry VII's accession." "On the historical side," the investigators continued, "there is at least a reasonable probability that the traditional story of the murder, as told by More, is in its main outlines true" (Tanner and Wright 1933).

After five days of on-site examination and description, the human bones were replaced in the urn with a statement on parchment detailing what had been done. The urn was resealed. To date no permission has been obtained for a reexamination that would afford an opportunity for DNA analysis, which was unknown in the 1930s.

The presence of animal bones in the urn, including duck, fish, chicken, rabbit, sheep, pig, and ox bones, can be explained, according to Tanner and Wright. The area over the secret burial hole, under or near stairs, accumulated quantities of kitchen waste in the 19 decades between the murders and the discovery of the remains. Such refuse would gradually have become submerged as the level of the soil built up. The workmen did not know in 1674 the significance of what they had found, smashed open the box, and threw the human bones onto a pile with other diggings. When Charles II ordered the royal remains removed from the tower, some of the human bones could no longer be located and some animal bones were inadvertently mixed in.

Alternative Explanations

It has sometimes been suggested that someone other than Richard III committed the murders, unbeknownst to him. The three names most often mentioned as alternatives to Richard are Henry VII, the Duke of Buckingham, and the Duke of Norfolk. Henry VII was not in any position to have the princes murdered

before he came to the throne in August 1485. The weight of contemporary authority points to the strong probability that the princes, last seen in the summer of 1483, were long dead by 1485. Buckingham, despite his ambitions, did not have a believable practical motive for the murders and would not have undertaken such a serious initiative without authorization from Richard III. Norfolk did have a financial motive for wanting the death of the youngest prince, Richard, Duke of York. The fact that young Richard had been declared illegitimate by parliament did not deprive him of the inheritance he stood to gain through the Mowbray Dukes of Norfolk, deriving from the early death of his child bride, Anne Mowbray. But it is not credible that Norfolk would have dared to commit such an act unless authorized to do so by Richard III.

Some of Richard III's supporters have hinted that the boys may have died of natural causes while incarcerated in the tower. Plague and the mysterious "sweating sickness" were both frequently fatal. The drinking water in the tower was considered unsanitary even by 15th-century standards. But if the princes had died a natural death, it would have been distinctly to Richard's advantage to display their bodies in public to allay suspicions that they had met with violence. A public funeral would have put the usurper in a more sympathetic light. He could have built support by exhibiting grief for his dead nephews, either real or fictitious. If an innocent explanation for the much-discussed disappearance of the little princes existed, Richard would have offered it. The fact that he made no such explanation strongly suggests that there was none to be made.

Conclusion

Contemporary authorities who discuss the probable date of the murders agree that the princes died sometime between the end of July 1483 and the end of September 1483. Most probably the slaughter occurred between August 7 and August 15. Richard III, at Warwick on his "progress" that week, sent Tyrell on a mission to the tower. If the dates are correct, Edward V was then just under 13 years old, and his brother Richard, Duke of York, was nearly 10. The skeletal evidence from the urn at Westminster Abbey shows that the two children were approximately 10 and 13 when they were killed. These dates and ages are not compatible with theories that Henry VII murdered the princes after he became king in 1485. The skeletons in the urn were not those of children aged 12 and 15.

Why, we may ask, if the princes were alive during Richard III's turbulent and dangerous reign, didn't he produce them for the public to see? In October 1483 a rebellion broke out in his kingdom in favor of Henry Tudor. Prominent Yorkists gave their support to Henry. Elizabeth Wydeville backed him with the understanding that, after winning the throne, he would marry her daughter (Edward V's sister), Elizabeth of York. The Wydevilles would not under any circumstances have supported Henry if the princes had still been alive at that time. Richard III

made no attempt to silence rumors about the death of the princes or to stabilize his position by showing that they were alive. His failure to take advantage of such an opportunity strongly suggests that the boys had already died. The accusations could have been easily dispelled if Richard III were innocent. When Lambert Simnel led a rebellion against Henry VII in 1487, impersonating the Earl of Warwick, Henry paraded the real Warwick through the streets of London to expose the imposture. But Richard could not publicly display the princes because they were dead, and he could not dispel the rumors because they were true.

In 1502 James Tyrell was awaiting execution following charges of conspiring with the Earl of Suffolk, a Yorkist pretender to the throne. Perhaps out of concern for their salvation, Tyrell and his old servant, John Dighton, made confessions in which they provided details about the death of the little princes. Their confessions were not made public by the Tudor authorities, but several versions found their way into circulation. Tyrell was beheaded, but Dighton, who was not arraigned, went free and openly told the story of the murders to anyone who would listen. Some historians have been puzzled about why Henry VII did not publish the confessions, which would have discredited Richard III and precluded future attempts by Yorkist pretenders claiming to be one of the vanished princes. Perkin Warbeck, for example, had led a rebellion against Henry in the late 1490s under the pretense that he was Richard, Duke of York, still alive after escaping the tower and living abroad in anonymity. But Henry VII was not inclined to remind the public of royal murders. He had himself recently arranged the beheading of the last Plantagenet male, the Earl of Warwick, on trumped up charges of conspiracy with Warbeck.

Richard III's apologists have often suggested that he had no motive for murdering the little princes, especially after parliament declared them illegitimate. But Henry VI had been restored to his throne after parliament had officially declared that he was a usurper. Edward IV had recovered his throne by an Act of Attainder. The Tudors showed little confidence in acts that disinherited their enemies and, instead, chose to ensure stability by means of judicial murder. As Niccolò Machiavelli wrote in his notorious tract on effective ("Machiavellian") governance, *The Prince*, a ruler who wants to survive must know how to do wrong. He "should not worry too much about conspiracies, as long as his people are devoted to him; but when they are hostile . . . he should fear everything and everybody" (1977).

Richard III did have compelling motives for killing the princes. His claim to the throne was weak. He was uncertain of his support. He suspected that Edward V, unjustly deposed, would serve as a convenient, invigorating rallying point for his enemies, for the sundry conspirators whom he constantly feared. He seemed to be in a constant state of nervous agitation throughout his reign and was often seen working a dagger, which he carried at his side, in and out of its sheath. In addition to motive, he had ample opportunity. The hapless children were secluded in a stronghold under the control of his men.

The exculpatory arguments put forward by Richard III's apologists ignore the full context of the situation between April 1483 and August 1485, dismiss inferences about Richard's character and state of mind that can be drawn from his known ruthless aggressions, and generally attempt to bathe him in an unreasonably positive light. The fact that his reputation suffered from attacks by Tudor detractors does mean that he was innocent. What actually happened to the little princes will always be subject to debate. But no one saw them alive after Richard III assumed the throne. Early accounts agree that they were murdered. There were many precedents in medieval England for dynastic murders, considered a wise precaution under circumstances analogous to those in which Richard III found himself in the summer of 1483. Considering the overall historical context and what we know about Richard III's character, it is highly likely that he instigated the murders of Edward V and his brother in the Tower of London. At the very least he was guilty of precipitating their deaths as a result of malicious neglect while they were incarcerated under his "protection."

As the proverb reminds us, "desperate times call for desperate measures," and perhaps it is less than fair to paint Richard III, as Shakespeare's *Richard III* does, as an intrinsically evil, vile schemer deliberately "determined to prove a villain." But the thoughts expressed by Shakespeare's Richard (in Act 5, Scene 3), on the night before Bosworth may come close to capturing the Machiavellian expediency of his reign and, by extension, the nature of his character:

> Conscience is but a word that cowards use,
> Devised at first to keep the strong in awe;
> Our strong arms be our conscience, swords our law. (Shakespeare 2008)

References and Further Reading

Commynes, Phillippe de. *Memoirs: The Reign of Louis XI, 1461–1483*. Harmondsworth, Middlesex, UK: Penguin, 1972.

Dockray, Keith. *Richard III: A Source Book*. Thrupp, Stroud, Gloucestershire, UK: Sutton, 1997.

Jenkins, Elizabeth. *The Princes in the Tower*. London: Hamish Hamilton, 1978.

Machiavelli, Niccolò. *The Prince*. New York: Norton Critical Editions, 1977.

Mancinus, Dominicus. *The Usurpation of Richard the Third*, 2nd ed. Edited and trans. by C. A. J. Armstrong. Oxford, UK: Clarendon Press, 1969.

More, Sir Thomas. *The History of King Richard III*. Bloomington: Indiana University Press, 2005.

Pronay, Nicholas, and John Cox, eds. *Crowland Chronicle Continuations: 1459–1486*. London: Alan Sutton Publishing for Richard III and Yorkist History Trust, 1986.

Ross, Charles. *Richard III*. Berkeley: University of California Press, 1981.

Shakespeare, William. *Richard III*. New York: Modern Library, 2008.

St. Aubyn, Giles. *The Year of Three Kings*. London: William Collins Sons, 1983.

Tanner, Lawrence E., and William Wright. "Recent Investigations Regarding the Fate of the Princes in the Tower." *Archaeologia* (1933): 1–26.

Thomas, A. H., and I. D. Thornley, eds. *Great Chronicle of London*. London: Guildhall Library, 1938.

Wood, Charles T. "Who Killed the Little Princes in the Tower." *Harvard Magazine* (January–February 1978): 3–40.

7

Columbus intentionally underestimated the circumference of Earth in order to get funding.

PRO   Talaat Shehata
CON   Joseph P. Byrne

PRO

The assertion here, that Christopher Columbus intentionally underestimated the distance he needed to travel to reach Asia in order to ensure that the Spanish Crown would fund his voyage, leads to other questions. How would Columbus "intentionally" underestimate the circumference of the Earth to help secure funding from the Portuguese and then the Spanish monarchy? What would justify the Spanish monarchs Ferdinand and Isabella's combined decision to fund a voyage based on a circumference for the Earth that was less than an actual, factual, and accepted reading of its measurements? The Crown had very mathematically and navigationally savvy authorities and committee members (the Talavera Commission, named after Hernando de Talavera, Queen Isabella's main confessor, who had been named to head the committee) to oversee the adventure. Over 500 years later, have the actions of the principal players borne fruit? Finally, what have been the positive and negative consequences of their individual and collective decisions? The answers to all of these questions will lead to the conclusion that it was in Columbus's interest to make his proposed voyage to Asia appear a simpler matter than he knew it to be.

First, to define a circumference, we need to brush up on our geometry principles. In the early Greek mathematician Euclid's *Elements*, he does not provide a clear definition as to a circumference's constructs. He does allude to its general form and outline. He equates the circumference with the shape of the circle. He emphasizes that when referring to a circumference, he was not discussing the measurement(s) *within* or diameter of the circle's interior; instead, he refers to it in the general all-encompassing form, outline, or shaped contour of a circle. During Columbus's days, and into the present, the circle is defined as encompassing its own circumference, length and all. A simple measure is that of the product pi times the diameter's length, or the product pi times twice the measurement of the radius. The numerical value often given pi is around 3.14159. The first letter for the word circumference in Greek is the pi symbol, as in *periphereia*.

141

Cristobal Colon (Christopher Columbus's Spanish name) wrote most of his works in his native Castilian tongue. Most of his accomplishments had been in the service of Castile. He had been self-taught in Portuguese, from his youth and into his early adulthood. There has never surfaced any solid historical evidence of him receiving or being grounded in any formal education. His local dialect was Ligurian, which never had a written format associated with it. So, as a result, when he came of age, his primary focus was on how to acquire and improve his verbal and written skills through his region's Castilian language. His command of the Italian language was nonexistent. He never wrote or carried on business in Italian, which, to a great extent is rather ironic, since in the United States and other parts of Europe, he's celebrated annually as a "native" son by most first-, second-, third-, and fourth-generation Italians.

To better understand Columbus the man, it is critical to better understand Columbus, the youth, who worked for years beside his father. It's very clear from the start that Columbus was, although emerging as the master seaman and maritime discoverer that he became in the 15th and early 16th centuries, first and foremost a first-class merchant. He helped his father in his wool and wine business. A shrewd businessman, he always had his priorities straight (i.e., how one consistently needed to secure and continue to increase the profit margins of the goods being bartered or traded as the never-ending bottom line). He was able to learn all the skills of a successful merchant and tradesman, at a very early and tender age, from his father. With his growing acquaintance with the rough and tumble of seafarers' lives and their often arduous and very challenging livelihoods, especially in the high seas, Columbus grew hardened in many ways, yet flexible in others. That, as the years rolled on, would define the complex historical and fascinating character that would emerge as the quintessential Christopher Columbus of folklore and legend.

What factors motivated the youthful Columbus to later become fixated on finding a new maritime passage to Japan, China, and Asia? In the never-ending quest for raw wool material and wine, young Columbus would, on behalf of his father and their business, take diverse voyages along the European and West African coasts. His travels were so widespread that they included the Genoese city of Chios, French Marseille, England, the Flanders, Iceland, Ireland, and Tunis in North Africa, Lagos, Nigeria, and deep into the African Gold Coast. It was following his voyage to the African Gold Coast that the idea began to germinate in his mind; that is, if the Portuguese and Europeans were able to sail so deep south along the West African Coast in quest of gold and treasures, then they should be as successful, possibly even more so, if they were to venture out into uncharted western oceans, which would lead them to a much shorter passageway (thereby eliminating western Europe's dependence on traditional road routes, caravans, bribes, and graft that since ancient times had grown to be much more exorbitant in their prices) to the riches that Europe had grown accustomed to acquiring from

the legendary Silk Road, over many centuries, from China. It would also illuminate encounters with the hostile tribal groups, not to mention their dreaded Muslim North African and West Asian neighbors, with whom King Ferdinand and Queen Isabella were engaged in a very violent insurrection to remove Muslim Moorish presence in Granada and Spain's other moderate to small cities.

China had for centuries established itself as the gold standard for all future trade. Columbus grew up in an expanding European naval and maritime period, which was dominated by Portuguese skill and expertise but still subordinate to the most sophisticated and technological seafaring global fleets of the Chinese Song Dynasties. This was specifically the case. With their mastery of the compass's use, the Chinese merchant marine and naval ships had established total dominance over the North Pacific Ocean lanes and much of the Indian Ocean lanes, which navigated around the Indian subcontinent and reached as far as Africa's eastern shores. It was seamless. They didn't need a Vasco de Gama to circumnavigate the Cape of Good Hope to reach the East and many of the Indonesian islands, where Europe would clamor to acquire its fine silks and spices. The Chinese were already there. They were the "East" that all future West European navigators, including Christopher Columbus, often stumbled over to figure *new* ways to get to their natural resources and finished goods. Most of the goods acquired by the Chinese in their eastern waters and lanes were much less expensive for the average Chinese merchant and consumer. That was so, even if they traveled to Spanish, Portuguese, or Venetian ports to acquire the goods they needed. Why should they have bothered? They wisely didn't. The "fire-sales" were regularly taking place closer to their own shores. All the Imperial government needed to do for China to maintain its reach, as they did with the well-known Ming Dynasty Admiral Zheng He, was project Chinese naval might as occurred in 1405 with gigantic displays of newly constructed armadas and ships, as it slowly traveled out of the Yangzi River and into the East China Sea. This occurred years *prior* to Christopher Columbus's birth, but it is very much a cultural recognition, if not appreciation, across synergetic cultural fault lines, that Columbus grew into adulthood surrounded by numerous intriguing tales of China, its military might and economic power, and its people's social and cultural context *within* the society. The Imperial government continued to regularly build, upgrade, and reinforce these naval and merchant marine vessels to make certain that the North Pacific region, with its seemingly never-ending archipelagos and islands, the Indian Ocean, and the East African coast remained stable and secure for continued profitable commercial acquisitions and gains for their nationals.

In his early 20s, Columbus, working as a deckhand on a return voyage from the African interior, in Lagos, Nigeria, experienced the trauma of having the ship he was working on sunk by a French mercenary ship. He would tread water for hours, until reaching Portugal's shores. Having no financial means of supporting himself, he headed to the capital city of Lisbon. Upon arrival in Lisbon, he was

fortunate to find a decent and caring Portuguese family to provide him with shelter and some living expenses, which he diligently worked off.

In 1479 the governor of the island of Porto Santo gave his daughter Dona Felipe Perestrello away in marriage to Christopher Columbus. The island that Dona Felipe Perestrello's father resided on was one of the group of islands identified with the Madeira Islands. After residing in Lisbon, then going back to reside with Dona's parents in Porto Santo, Columbus and his wife were blessed with the birth of their only child, whom they named Diego. With his new responsibilities, Columbus continued to dedicate his efforts to becoming a more skillful and profitable merchant. With his continuing and regular visits to the gulf coast of Africa and other more distant parts of the West African coastlines, Columbus's obsession with finding a new passageway west to Japan and China grew much stronger. As a merchant, first and foremost, he contemplated the riches in gold, silver, other goods, and vast real estate that might lie out there, beyond the reach of all Europeans. If only he, through the aid of the Portuguese Crown or some other rich and strong benefactor(s), could have a few trips funded across the Atlantic Ocean for such a specific purpose. He would call it the "Enterprise of the Indies."

According to his calculations, it was too good an opportunity to be passed up by any of the competing European monarchs. He would attempt to appeal to each of their growing political, economic, and naval needs and self-interest during this new Age of Discovery and Exploration, increasing nationalism, and gradually evolving imperialism, to seize and gain more control over a larger global empire. Being a devout Christian, he was described by his illegitimate son Ferdinand, who after Columbus's death, would later be recognized by scholars as his sole and most accurate biographer, as being "so strict in matters of religion, he might have been taken for a member of a religious order." (McCrank 1993: 236) He would use that shared common religious theme and sentiment for the conversion of native people he might encounter in his "Enterprise of the Indies" into the Christian fold as yet another means by which he might gain funding from the European "Christian" monarchs that he might encounter in making the greatest sales pitch in his long career as a naval sea merchant.

So, even decades *prior* to the arrival of the first Spanish missionaries in the Caribbean, Central and South America's coastlines, and that of the earliest Calvinist pioneers, who settled the East Coast shores of the "new" North American continent, Columbus had set the tone for the Bible-thumping proselytizing Christianity that became a regular feature of North and South America's religious history. Columbus would infuse a mystical quality of his very practical venture into one of capitalism's earliest experiments on a grand scale of mass accumulation and consumption of the silver, gold, diamond, pearl, and spice finds and treasures, with the newly introduced slave system into the Atlantic economic and political systems of trade, on the three continents of Africa, South America, and North America. His adventure would result in a regular exchange of goods, the

conflicts and synergy between the Old and New Worlds in the areas of religion, vast technological differences, the insidious onslaught and decimation of a large percentage of the Native American populations by different diseases unknown earlier to them, food production and regular warfare within Native American tribal groups and against all other foreign (as the Spanish, British, and French) intruders, and finally, the gradual emergence of regular episodes of violence and resistance against enslavement and exploitation. They would all be a direct outcome of the goals that Columbus had set for himself, Spain, and the rest of Europe. He strongly believed, or so he had all others convinced, that he was divinely inspired, and that gave this mission the most steadfast sense of purpose. Now, the trick was to convince the European monarchs of the divinity of such a mission, so that they and the appointed royal committees that would look into Columbus's proposal would equally share in his passion, zeal, and purpose. It would be a hard undertaking, if viewed by any other measure. But the most significant historical development of these meetings and their future consequences was that it would be a pivotal moment in world history, where the ancient and medieval Mediterranean-dominated world, along with the medieval and postmedieval Indian Ocean and Pacific Ocean Chinese-dominated world, would never again be the same. They would recede and give way to a new global world order, dominated by West Europeans, for the first time in their history overtaking all their global competitors within a few brief centuries, and show both its constructive and negative qualities in the formulation of a postmedieval and free-enterprise, scientific, and more rational modern world.

Therefore, doing the decent and loyal thing as a native son, Columbus took his offer to the Portuguese King John II. After having a royal maritime committee scrupulously give his proposal a quick once over, the shrewd maritime experts that they were refused Columbus's construing of the figures. That is, he suggested that a long voyage to Japan from the Canary Islands would be a mere 2,400 nautical miles. Instead, the committee elders were quite wise to the fact that such a journey would never be less than around 10,600 nautical miles. He clearly had consciously underestimated the circumference of the Earth, and specifically Asia's distance from the point of origin in the Canaries. What would have caused Columbus to fudge the figures? First, Columbus's overwhelming ambitions appeared to dictate what bogus information he would be willing to put across to the committee members. Second, there was the fear that if he provided them with the accurate figures, which he and they already knew as longtime sea navigators, he'd have his funding refused because it would be too costly and equally hazardous for his ships and seamen to make such a potentially deadly voyage into the unknown. For Columbus, this second option would have been a glaring lose-lose situation. So, he rationalized, a little fudging of the numbers and facts wouldn't hurt the overall venture, especially if the committee members fell for the sleight-of-hand ploy. They were too wise and tried for a scheme such as the one that

Portrait said to be of Christopher Columbus by Sebastiano del Piombo, 1519. (Library of Congress)

Columbus sought to lob by them. Finally, if the pitch didn't work with them, he knew he always had a handful of other European monarchs who would be more than willing to consider his grand proposals. Their gluttonous capacity for more and more wealth and power was incalculable. As a tradesman and merchant, Columbus had grown to be a very shrewd judge of human character.

In 1485, Dona Felipe had passed away. Christopher was left with Diego, his only son by her. After being refused by the Portuguese, Christopher and Diego, hand in hand, traveled to Spain for the next sales pitch. During his stay in Cordoba, he met Beatriz Enriquez de Harana, who would later become his mistress and the mother of his second son, Ferdinand. Columbus had Antonio de Marchena, a very learned Franciscan friar, intercede for him with King Ferdinand and Queen Isabella. Despite the fact that Hernando de Talavera was Isabella's confessor and head of the commission to which Columbus's plans were to be presented, members of the commission, like their Portuguese counterparts, were equally suspicious of Columbus's loose ways with fudging the numbers and facts. They refused his proposal. Ferdinand and Isabella consoled Columbus's rejection by the royal commission and continued to pay him for his services to the Crown. Ferdinand and Isabella had already had their minds set on taking that gamble, with the finances they knew they would need, in order to consolidate their power in Spain after the expulsion of the Muslim Moors; but also just as important, because of their jockeying for position of supremacy, among fellow European monarchs, with an expanding Spanish imperial reach and presence in the Old and New Worlds. The only matter where any serious disagreement arose between Columbus and the Spanish monarchs was the issue of him being granted the title of "Admiral of the Ocean Sea, Viceroy, and Governor." There were also contracts and ultimate capitulation by the Spanish monarchs to Columbus (in what now is historically known as the Capitulations of Santa Fe) on matters of real estate ownership of any new property he might lay claim to for the Spanish Crown, himself, and all his future descendants. Finally, Columbus asked that in all new land "discovered," Columbus and his descendants

would automatically acquire 10 percent of all items and goods traded and pro-
duced; along with 10 percent of all earnings from the gold, silver, gems, pearls,
and all other valuable treasures discovered. Ferdinand in particular was rather
hesitant to provide Columbus with all that he had wished for and ultimately
demanded from the Spanish Crown. This would later emerge as a very serious
sticking point between Ferdinand and Columbus; especially, once his hunches
proved accurate regarding the gold, silver, treasures, and riches of the New
World. But, Ferdinand and Isabella, being very competitive cousins of the Por-
tuguese King Joco II-Joco (João) the "Perfect," were not to be later outmaneu-
vered by him or any of their other cousins, within the *intimate* European royal
family disputes that often surfaced. They were determined to acquire more
wealth and power to help in the financing of their future anticipated wars of
conquest and overall domestic unification.

A comical development after the return of Columbus from his first trip to
the New World was when he was invited by the Portuguese King Joco II-Joco to
stop by Lisbon on his way back to Spain to be received as a conquering hero.
King Joco, despite turning Columbus down for his bold ideas about a voyage to
the New World, with the help of his appointed royal maritime committees, was
not about to be overshadowed by his cousins Ferdinand and Isabella's moment
of triumph, after having agreed to fund Columbus's voyage. It was useful that
King Joco II's great-uncle was Henry the Navigator. He had become famous for
subduing the Moors along the Moroccan shores and had gone deep into West
and Central Africa to attempt to lay claim and control over the very important
gold trade that cut through numerous West African states and tribal grounds. So,
King Joco, in his attempt to later "rope in" Columbus and the Spanish mon-
archs' bounty, shrewdly embraced Columbus's accomplishment as brilliant and
beyond reproach, showering him with effusive phrases at each opportunity that
presented itself when they mingled with the Portuguese. King Joco also decided
to deceitfully, or in chutzpa-like fashion, like a fox in a hen house, lay claim to
any treasures brought back from the New World, since in the Eighth Article of
the 1479 Treaty of Alcagovas, it was stipulated that all territories parallel and
directly opposite to the West African coast, and south of the Canary Islands,
would be considered Portugal's. Ferdinand and Isabella were seething at the sug-
gestion, and they immediately outmaneuvered King Joco by going to a higher
authority, the pope. They based their argument on an old statute, from the times
of the Crusades. The statute basically affirmed the right of good "Christian" sol-
diers and authorities to seize all heathen property, so that the Catholic faith could
continue to grow and expand. Pope Alexander VI, famous historically for many
of his antireligious antics (having seven children out of wedlock and having pur-
chased the papacy instead of being appointed to the office by his hard work,
contributions, and piety), thought it was a no-brainer—he would side with Ferdi-
nand and Isabella. It also helped that he was a Spaniard, who had always had

very strong ties to his homeland. Also, one final and very important point, once the pope publicly made his decision, there was no grievance process allowed for anyone to dispute his *holy* edict. The game was over for King Joco and the Portuguese before it had even started.

But, to not get too far ahead of ourselves, let's rejoin the Spanish Royal Commission's negative report and Ferdinand's early reticence toward Columbus's demands, *prior* to gaining permission and funding from Ferdinand and Isabella for his first voyage. Columbus found himself, once again, faced with yet another rejection similar to the one he received in Lisbon. Disappointed, but yet unperturbed, he took his young son Diego, got on a mule, and proceeded on his journey to Seville. At Seville, Columbus intended to take a ship to France to continue his sales pitch to the French monarch, Charles VIII. One of the most forgotten, but equally most pivotal personalities in world history, Isabella's court official Luis De Santangel was so disappointed by the Spanish monarchs and Royal Commission's decision that he pleaded with Queen Isabella to rescind her and Ferdinand's decision. He felt that Spain would emerge from this entire affair as either the greatest loser if they didn't agree to Columbus's terms, or the greatest winners, as at least two and a half centuries of future Spanish gold, silver, and rare treasure and trade acquisitions and Atlantic trade dominance would later attest. Isabella, finally swayed by Santangel's passionate, persuasive tactics, had a court courier intercept Columbus a few miles prior to his arrival at Granada, on his later trek to Seville. The rest is history.

All matters being at least momentarily resolved by the principal parties; Columbus got three caravels, the *Niña,* the *Pinta*, and the *Santa Maria*. The *Niña* and the *Pinta* were given to Vicente Yáñez Pinzón and Martin Alonso Pinzón, respectively, to command. Columbus assumed command for the *Santa Maria*. With the voyage costing in the vicinity of $14,000, they set sail on the morning of August 3, 1492, from the port of Palos, Spain, for their unknown voyage to the New World. Columbus's primary fixation, which remained with him throughout his other three voyages to the New World, was how he and his men could penetrate into the interior of "Asia," instead of regularly merely "touching base" at its fringes. When his men later landed on the shores of Cuba, he would send appointed emissaries, failing of course, to the "Great Khan" of China. He was completely clueless as to the fact that he had touched upon a whole new continent. The only inkling of the fact that he was onto something much grander than he had ever anticipated was when he used scripture to buttress his "Enterprise of the Indies" plan. He referred to the Orinoco, a river he came upon in South America, as being of Paradise's four rivers. He also declared on his third voyage, which began from Spanish shores on May 30, 1498, that upon arriving at the Gulf of Praia, located on Venezuela's coast, he believed "that this is a very great continent, until today unknown." As for Columbus, he had not internalized and digested the stark reality of a new continent; but, instead, with his

---

Columbus's First Impressions of America

When Christopher Columbus first arrived in the Caribbean, his descriptions of the people he found there were recorded for posterity by a man who would later be noted for his defense of the Americas' indigenous populations, Bartolomé de las Casas:

> As I saw that they were very friendly to us, and perceived that they could be much more easily converted to our holy faith by gentle means than by force, I presented them with some red caps, and strings of beads to wear upon the neck, and many other trifles of small value, wherewith they were much delighted, and became wonderfully attached to us. Afterwards they came swimming to the boats, bringing parrots, balls of cotton thread, javelins and many other things which they exchanged for articles we gave them, such as glass beads, and hawk's bells; which trade was carried on with the utmost good will. But they seemed on the whole to me, to be a very poor people. They all go completely naked, even the women, though I saw but one girl. All whom I saw, were young, not above thirty years of age, well made, with fine shapes and faces; their hair short, and coarse like that of a horse's tail, combed toward the forehead, except a small portion which they suffer to hang down behind, and never cut. . . .
>
> Weapons they have none, nor are acquainted with them, for I showed them swords which they grasped by the blades, and cut themselves through ignorance. They have no iron, their javelins being without it, and nothing more than sticks, though some have fish-bones or other things at the ends. They are all of a good size and stature, and handsomely formed. I saw some with scars of wounds upon their bodies, and demanded by signs the cause of them; they answered me in the same way, that there came people from the other islands in the neighborhood who endeavored to make prisoners of them, and they defended themselves. I thought then, and still believe, that these were from the continent. It appears to me, that the people are ingenious, and would be good servants; and I am of opinion that they would very readily become Christians, as they appear to have no religion.

Source: Christopher Columbus and Bartolomé de las Casas. *Personal Narrative of the First Voyage of Columbus to America.* Philadelphia: Thomas B. Wait and Son, 1827.

---

passionate and religious predilections, he was convinced that he had finally come upon the Garden of Eden, as described in the Bible.

Having landed safely in what is present-day San Salvador, October 12, 1492, on his first voyage, Columbus in his already self-imposed state of denial in many matters religious, which often bled as we've seen into his professional duties and discoveries, called out to his "indios" receiving native population party. The reason for that was that he was convinced he had come upon the East Indies. Many of the wiser men surrounding him, and mapmakers in Spain, immediately recognized the errors of his ways and that he had come upon the

Arawaks and the group of neighboring islands they had inhabited for generations. But, despite that, the first voyage and his return to Spain in March 14, 1493, was his finest hour. The Spanish populace and the royal family celebrated Columbus and his men's accomplishments as returned conquering heroes. But, by the second voyage, which began September 5, 1493, and ended March 10, 1496, the novelty of his voyage and accomplishments began to fade. *Prior* to the end of his second voyage, he had to rush back to Spain to try to defend the honor of his brothers, who had been accused by the authorities with gross negligence in governance and cruelty to the native inhabitants. By the third voyage, Ferdinand began to exert more of his "royal" authority and appointed a new governor in the island of Hispaniola, in the New World, by all indications, cutting into Columbus's authority as the "supreme" viceroy and governor of the territory. The Spanish Crown's newly appointed governor, Francisco de Bobadillaas, shortly after assuming command of Hispaniola, put Columbus and his brothers in chains and sent them back to Spain. Upon arrival, Ferdinand and Isabella publicly showed their deep displeasure at what and how their "Admiral of the Ocean Sea, Viceroy, and Governor" had been treated. Privately, Ferdinand and his surrounding sycophants walked away with silently shared smiles, proud of the fact that they had cut the "Admiral of the Ocean Sea" down to size. Now they could gradually begin to deny him and his descendants the legal share of the obscene riches that would come to Spain and the monarchy from the New World. This would also include the back pay for Columbus's officers and sea crew, which until he died in Valladolid, Spain, on May 20, 1506, he continued to fight the Crown and Spanish government to reimburse them with.

Whether Columbus intentionally underestimated the circumference of the Earth in order to get funding is historically a moot point. The reality is that if Columbus had not had the backing of any one of the European sovereigns of his day, he never would have brought about the historical consciousness of the globe to attain its current level of realization. Columbus's endeavors helped set in motion forces that even he had never contemplated. Instead of remaining in a frozen periphery status to the mainstream of global events, the American continent, with ongoing western expansion from Europe and other nations and into its deepest hinder lands, would in a few short centuries be poised to define the *innate* essence of the mainstream.

The new political, economic, cultural, and social experimentation that would regularly occur in these new settled shores—such as the American Revolution in 1776 and the Slave Rebellion in Haiti, and the revolutions of the early 19th century in South America, with revolutionaries as Simon Bolivar and others—would impact Europe and, in time, most nations of the world. No matter what most scholars or readers might think of Columbus's "fudging" of the truth when it came to the Earth's circumference, it might be concluded that with 500 years of hindsight for us to mine and benefit from, it would not be exercising too much stretch to

one's historical imagination if one concluded that instead of calling Columbus "divine," as his sailors and officers often did in recognition of his mastery of the oceans, we could call what his bold actions brought forth as having a transformative impact on human history, both for good and bad. We've also been touched in our daily postmodern lives by what Columbus accomplished. One could argue that the world, with many of the negative and positives weighed alongside each other, is a much better world for the tenacity and conviction that Columbus brought to it. He left it a much more vibrant and competitive place than he had found it. Hope for a better life in the New World became, and continues to be, the *new* and undiminished *mantra*. Therefore, in the attempt to address his fudging of the facts about Earth's circumference, maybe future generations can humanely give him a little slack. He clearly seems to have earned it.

References and Further Reading

Anderson, E. W. *The Principles of Navigation*. New York: American Elsevier Pub. Co., 1966.

Andrews, Charles M. "The Acts of Trade." In *Cambridge History of the British Empire*, Vol. 1. Edited by J. Holland Rose et al. New York: Cambridge University Press: 1929.

Bigelow, Bill, and Hans Koning. *Columbus: His Enterprise: Exploding the Myth*. New York: Monthly Review Press, 1992.

Dugard, Martin. *Last Voyage of Columbus: Being the Epic Tale of the Great Captain's Fourth Expedition, Including Accounts of Mutiny, Shipwreck, and Discovery*. New York: Little, Brown, 2006.

Fredrick J. Pohl, and L. B. Loeb. "Americo Vespucci—Pioneer Celo-Navigator." In *United States Naval Institute Proceedings* (Vol. 83, No. 4, pp. 396–403). Annapolis: United States Naval Academy, 1957.

Horwitz, Tony. *A Voyage Long and Strange: Rediscovering the New World*. New York: Holt, 2008.

Kaplan, Lawrence S. *Colonies into Nation: American Diplomacy, 1763–1801*. New York: Macmillan, 1972.

McCrank, Lawrence J. *Discovery in the Archives of Spain and Portugal: Quincentenary Essays, 1492–1992*. Binghamton, NY: Haworth Press, 1993.

McNeill, William H. *The Rise of the West*. Chicago: University of Chicago Press, 1963.

Morison, Samuel Eliot, ed. and trans. *Admiral of the Ocean Sea: A Life of Christopher Columbus*. New York: Little, Brown, 1991.

Morison, Samuel Eliot, ed. and trans. *Journals and Other Documents on the Life and Voyages of Christopher Columbus*. New York: Little, Brown, 1964.

Nef, John U. *War and Human Progress*. Cambridge, MA: Harvard University Press, 1950.

Palmer, Robert R. *The Age of Democratic Revolution*, Vol. 1. Princeton, NJ: Princeton University Press, 1959.

Sauer, Carl Ortwin. *The Early Spanish Main*. Berkeley: University of California Press, 1966.

Vignaud, Henri. *Histoire critique de la grande enterprise De Christophe Colomb*, 2 vols. Paris: Welter, 1911.

Young, Filson. *Christopher Columbus*. New York: IndyPublish, 2009.

CON

In the quincentennial year of 1992, the world took a long, hard look at Christopher Columbus and his "Enterprise of the Indies." Historians began the task of creating new and definitive editions and translations of the documents related to him; they contextualized him by reexamining the Old World from which he sailed and the New World that he opened to European influence and conquest, for good and for ill. The year saw the publication or reprinting of dozens of books and articles outlining Columbus's personal life, dreams, career, and achievements in a generally favorable light. On the other hand, some historians and others sympathetic to the tragic history of Native and African peoples in the post-Columbian Americas emphasized the brutality of the Iberian colonization and exploitation, the beginnings of the transatlantic slave trade, and Columbus's own flaws as a human being and administrator. Rather than celebrating the "Admiral of the Ocean Sea," they decried him as slaver, charlatan, and father of Amerindian genocide. Rather than "noble visionary," he was dubbed "obsessed," "self-righteous," "racist," "arrogant," and "incapable of telling the truth."

The Controversy over What Columbus Believed

One element in this controversy was the origin of his "Enterprise" in his own mind. By at least the mid-1480s this Genoese merchant and sometime sailor had formulated a plan for crossing the Atlantic Ocean in order to reach what many believed would be the coastal waters and eastern archipelagos of the Asian continent. When he tried to sell his idea to King João II of Portugal in 1484, he was sent away. Relocating to Castile, he soon began pitching his vision to the Spanish monarchs, Ferdinand and Isabella. As with João, they were advised by their experts against taking the risks that supporting Columbus's voyage would entail. Instead of dismissing him, however, they held Columbus at bay until 1492. As a frustrated Columbus was on the verge of visiting the French court to make the

same proposal, Isabella relented and provided the required men, ships, and supplies. The rest is history.

In fact, Columbus's calculations and assumptions concerning such a voyage were quite incorrect; no one argues about that. As with all things Columbus, however, his denigrators claim that Columbus was a charlatan who knew quite well that the geographic elements of his proposal were not only flawed, but also wrong. These elements rested on two fundamental assumptions: (1) the distance of circumference of the Earth at the equator; and (2) the portion (or percentage) of that circumference covered by the Atlantic Ocean between Europe and Asia, which Columbus would have to cross (no one having even dreamed of the existence of the Americas). The latter issue was directly related to the distance, as a percentage of the circumference, across the Eurasian landmass from Spain eastward to farthest China: if, for example, Eurasia covered 50 percent of the distance, then the Atlantic covered the other 50 percent. For Columbus's plan to work, the distance across the open Atlantic had to be minimal, which meant that (1) the circumference of the Earth (no educated person at that time thought the Earth was anything but more or less spherical) had to be as small as possible; and (2) the extent of the Eurasian landmass had to be as large as possible. Clearly, both the Portuguese and Spanish court experts who denied Columbus their support believed that the distance across the Atlantic to Asia was too great to cross, at least with ships and navigational technologies as they were. With no America in the picture, they were certainly correct.

Of course, if there were areas of dry land in the ocean between Europe and Asia, then a ship or squadron of ships could island hop across, landing as needed for food, water, rest, and repair. Tales and theories abounded about such isles. After all, Europe had its British Isles and Azores, and Africa its Canaries, Madeiras, and Cape Verde Islands, so why discount stories of Cipangu (Japan) and literally thousands of islands mentioned by the 13th-century Venetian merchant and traveler Marco Polo? Why discount the mid-Atlantic islands supposedly reached by the Irish monk Saint Brendan on his (mythical) Atlantic voyage? Why discount the traditions of Antillia, the Hesperides, the Amazons, and Brazil? Columbus, and many others, did not discount the possibility of such islands between Europe and Asia, in fact he came to count on them existing.

So, the world that Columbus presented to the Iberian monarchs was small, its Eurasian landmass quite extensive, and its eastern waters amply provided with islands—richly endowed islands, he assured them—that stretched well into the western Atlantic. Did he actually believe this to be the case, or was he knowingly presenting a false picture in order to convince the rulers to accept and fund his venture? The great majority of historians from his earliest biographers to most writing today have accepted that, while obsessive, Columbus fully accepted his "crossable Atlantic" theory. Revisionists, who generally hate Columbus and what followed in his wake, try to argue that he could not have believed in a "crossable

Atlantic," and that his proposal was filled with carefully orchestrated lies. The problems in sorting this out are that (1) we do not have the texts of Columbus's proposals to either throne; (2) nor do we have the reports of the experts arguing against the proposals. Moreover, we have only inferential evidence for what Columbus believed and when he believed it. Even if Columbus had clearly outlined for posterity his beliefs and their rationale, need we not be skeptical? The controversy, then, has to focus on the question, could Columbus have believed in a "crossable Atlantic"? This takes us into "Columbus's worlds," as some have called it: what he experienced, read, and heard before he made his proposals that convinced him that he could make the voyage that none before him ever had.

Columbus's Sources

Our knowledge of the life and beliefs of Christopher Columbus comes from his own writings and the biographical material provided by his son Fernán (or Ferdinand) Columbus (or Hispanized as Colón) and Columbus's friend the famous champion of the natives of New Spain and bishop of Chiapas, Bertolomé de las Casas. Columbus's own writings that shed light on this controversy include his records of his first voyage, in which he reflected on his beliefs, and marginal notes—glosses or postils—in several printed books that comprised his personal library. These were preserved by his son and exist today as part of the *Biblioteca colombina* (Columbian Library) in the Chapter Library of the Cathedral of Seville, Spain. In his biography of his father, Fernán clearly states that Columbus's vision was shaped by three factors: his own experiences at sea, his conversations with other mariners, and his reading of classical and medieval sources on what was known as "cosmography," or the physical structure of the known world. Though Fernán privileges Columbus's reading as the principal impetus, seeking to paint his father as a scholar, most modern historians believe the "Enterprise" began with his own experiences.

Christopher Columbus was born in 1451 into a family of weavers and cloth merchants in Genoa, one of Italy's ancient maritime cities. Situated on a narrow strip of coastline, Genoa had had a long history of naval and mercantile success, but its glory days were past when Christopher began his career. Fernán claims that Columbus first went to sea when he was 14, probably accompanying family goods if true. In his early years around docks and warehouses he no doubt picked up much lore that derived from centuries of Mediterranean voyaging and more recent expeditions into the Atlantic to reach English, Flemish, and Scandinavian markets. We only have glimpses of his own experiences, but it seems that he sailed as far east as the Aegean island of Chios, and, perhaps in 1470, traveled the Atlantic to England or Flanders on family business. A turning point came in August 1476, when the 25-year-old Columbus was shipwrecked during a French ambush at sea off the coast of Portugal. He came ashore near Lisbon,

Portugal's capital, and from that time never again gave much thought to the Mediterranean. He claimed that the following February found him sailing northwest beyond the famed island of "Tile" (classical Thule; probably Iceland), which ancient writers considered the known world's westernmost point. He noted that is was much farther west than the second-century Roman geographer Ptolemy—who probably had Shetland Island in mind—had indicated.

By the early 1480s Columbus had married the Portuguese noblewoman Felipa Moniz e Perestrella, whose father was governor of the Madeiras Islands. Columbus presumably spent no little time among the islands when not in Lisbon. He also spent time in the Spanish-controlled Canary Islands, studying the winds and currents and absorbing local lore. At some point in the early 1480s he sailed down the African coast as far as the new Portuguese settlement at São Jorge da Mina, south of Cape Bojador on the Guinea coast. Significantly, this was below the equator and populated, and thus dispelled a venerable myth that humans could not survive below the equator. In addition to providing important information about winds and currents, these ocean voyages also allowed Columbus to test his theories about the length of a degree, or 1/360 of the Earth's circumference at the equator, or so Columbus recorded. This determination was crucial since its value determined the Earth's circumference.

During his time in the islands he is also thought to have seen, and certainly he heard of, flotsam (debris) that washed up on island shores from points west. Human-made wooden artifacts, the remains of unfamiliar trees, and even human corpses were known to float ashore on islands from Ireland to the Cape Verdes after storms that rode in from the west. While this evidence did not prove that western isles (or Asian ones) were within sailing distance, they certainly suggested it, and further implied that the currents and winds that propelled junk could propel sailing ships.

The vastness of the Atlantic had spawned speculation about Atlantic isles since antiquity. The 15th century had seen the rediscovery of the Madeira archipelago (Romans had called them the Purple Islands), the discovery of the Azores and Cape Verde islands, and the colonization of the Canaries; the notion that there were none left to be discovered was virtually illogical. Both native and European mariners claimed to have sighted distant shorelines while deep in the Atlantic, seemingly confirming the existence of islands long imagined. One tradition has it that a dying Portuguese ship's pilot (the "unnamed pilot") conveyed to Columbus the details of his unexpected landing on the legendary island of Antillia, the storm-tossed return from which killed his entire crew. Some suspicious historians even came to believe that Columbus himself had reached the Atlantic's far shores, and only wanted the royal support to lay claim to them (a notion long since discarded). If mariners' tales of nearby islands were false but titillating, the details they provided of open ocean winds and currents were crucial to Columbus's calculations.

But Columbus was not alone in treating rumors of undiscovered islands seriously. Marco Polo had claimed there were 1,378, and the Catalán *Atlas* of 1375 indicated there were 7,548. In fact, kings Afonso V and João II of Portugal had made charters for explorers venturing into the Atlantic long before Columbus ever broached the subject. On February 19, 1462, Afonso made a grant to João Vogado of the islands of Lovo and Capraria, supposedly landed on by the sixth-century Irish monk Saint Brendan. A dozen years later he granted a charter to Fernão Teles to all islands he should discover *up to* Antillia, as long as they were not in the vicinity of the Guinea coast. On June 30, 1484, the year João II sent Columbus away for his unreasonable proposal, he granted a royal charter to Fernão Domingos do Arcoan for all lands he *intended* to discover in the Atlantic. And on March 3, 1486, two years after Columbus left Portugal for Castile, João granted the mythical island of Antillia to Ferdinand van Olm (Fernão Dulmo, a Flemish resident of the Azores), based on his proposed voyage of discovery. The king also promised royal military aid to subjugate the populace if needed. Perhaps significantly, none of these grants indicated mentioned Asia, and this may indicate the main reason for Columbus's lack of success in Lisbon. The grant to Dulmo was to be the last before 1492, what with the successful voyage around Africa of Portuguese explorer Bartolomeu Dias (1486–1488). This last, and earth-shattering voyage, may well have been prompted by the failure of José Vizinho to cross the African continent to the Indian Ocean, a fact he reported to the Portuguese court in 1485. This was also a fact that supported Columbus's belief that west-east landmass distances were far greater than generally supposed, and thus the distance across the Atlantic far less.

A Matter of Degrees

What of the circumference of Earth and the distance in miles from Portugal, or its islands, to Cathay (China) or its eastern outlier, Cipangu? To answer these questions 15th-century Europe looked backward to the writings of classical and medieval writers on "cosmography." This was fully in line with the intellectual pursuits we label Renaissance or humanistic and is behind Fernán's claims that his father drew his initial inspiration from the Greeks and Romans, rather than the salty sea-dogs of his own day. Modern biographers generally accept that Columbus's reading helped confirm his ideas rather than shape them, however, and that it was vital in bolstering his proposals to the intellectuals at court. We have to begin with the texts that Columbus actually owned and annotated and that are now in the *Biblioteca colombina*. We know their publication dates, and thus can provide a *terminus post quem*, or date *before* which Columbus could *not* have owned them. But does this mean that Columbus could not have read other, earlier copies of the same works, perhaps even in manuscript form (Columbus was three when Gutenberg invented the printing press)? It does not, though no evidence

Just How Long Is a Mile?

There have been many definitions of a 'mile' of length, which have been used in different parts of the world at different times in history. Perhaps the earliest definition was the Roman mile, or *mille*, which was supposed to be the distance that a legion could march in 1,000 paces. The Romans placed milestones along their numerous roads at approximately 1,520-meter distances from one another. Britain, Scotland, Ireland, and France had mile measurements during the Middle Ages, which ranged from approximately 1,600 meters to over 2,000 meters. The modern mile was established in 1592 by the British Parliament as measuring eight furlongs, which works out to approximately 1,609 meters. With all of this uncertainty, it should not be a surprise that Columbus might have been confused!

supports either side of this debate. And so most modern biographers assume he first read the copies he owned.

The *Geography* of Claudius Ptolemy was considered the most authoritative "cosmographic" work of the ancient world. It was written about 148 CE, first translated into Latin in 1406 in Florence by Jacobus Angelus, and first printed in 1475, with four additional editions by 1492. Columbus's copy was printed in Vicenza in 1478 and has none of the maps of other editions. Ptolemy posited that the Eurasian landmass was continuous, that islands indeed peppered the sea between its two ends, but that it covered only one half of the globe (177 degrees), leaving a vast expanse of ocean. But how vast? This depended on the size of a degree of longitude. The ancient Greeks measured in "stades" (hence our word stadium), eight of which made up a "mile" if they were Olympian, but only 7.5 of which made up a mile if they were Phileterian, the standard used during Roman times. The Greek scientist and philosopher Aristotle posited that the Earth's circumference was 400,000 stades, and that a degree consisted of 1,111.11 stades, or 138.9 Roman miles, making the Earth's circumference some 50,000 miles. Either the famous scientist Archimedes or Dicaearchus of Messana (the source is unclear) ventured that the real number was 300,000 stades. The Greco-Roman astronomer Eratosthenes, who actually bothered to carry out a fascinating experiment to determine the figure scientifically, found it to be 250,000 stades, though he fudged it to 252,000 to get a degree equal to an even 700 stades (252,000 ÷ 360), or 87.5 Roman miles. By Ptolemy's time, the number of miles comprising a degree had been lowered to 66.66 miles (Posidonius) and 62.5 miles (Strabo), which would have made the total distance 22,500 Roman miles, the figure accepted by Ptolemy and his contemporary Marinus of Tyre. Ptolemy's work was well known to medieval Islamic scholars, including the ninth-century expert on Ptolemy, Alfraganus (al-Farghani), who witnessed an Erastothenes-like experiment in the Syrian desert by one of the Caliph's astronomers and reported it in his *Elements of Astronomy*. This purportedly proved the
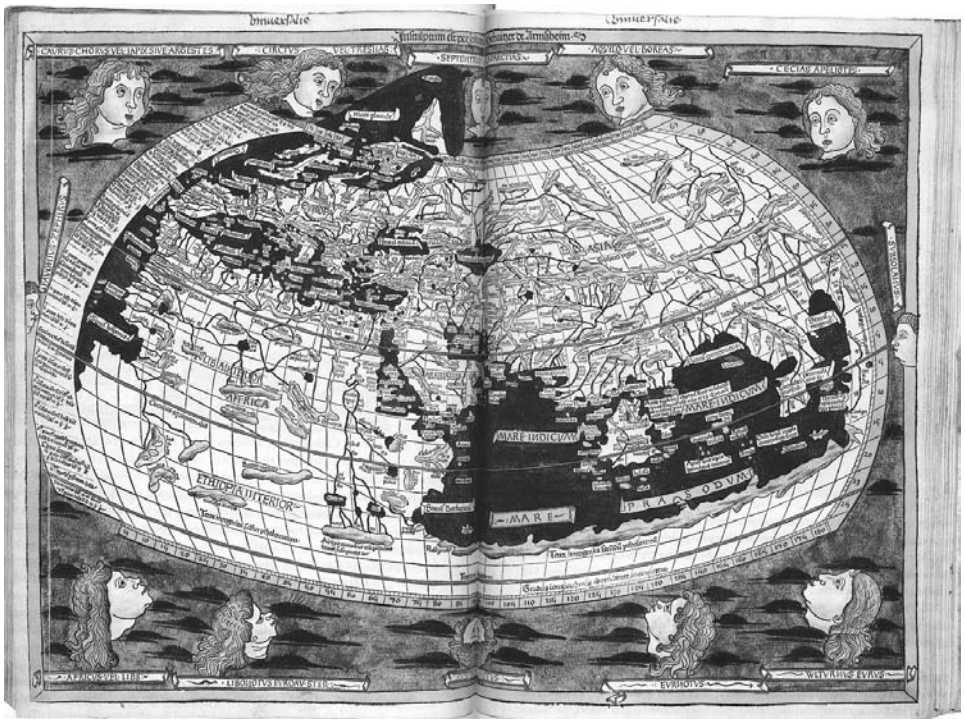
Ptolemy's map of the world, printed at Ulm in 1482. (Library of Congress)

degree to be 56.666 Roman miles, which would make the total distance 20,400 miles, a far cry from Aristotle's 50,000. The problem was that Alfraganus's mile was not the same as the Roman mile, in fact it was rather longer: Roman mile was 1,481.75 meters, Arab mile was 1,973.5 meters (the modern mile is 1,609.34 meters; the equatorial circumference is in fact 24,901.5 modern miles). This was a fact that may have been known by the scholars of the day, but Columbus was not among them: a mile was a mile. He had found the figures that supported his adventure: a degree at the equator of 56.66 miles and a total circumference of 20,400 miles. Interestingly, Columbus claimed to have confirmed Alfraganus's 56.66-mile degree during his coasting of Guinea.

Now Columbus had probably not read Alfraganus, even in the 12th-century Latin translation of John of Spain and Gerhard of Cremona. The late 14th-century cardinal and University of Paris theologian Pierre d'Ailly, however, did, and recounted the whole controversy, in his *Imago mundi* (Description of the World) of 1410. D'Ailly's work was a compendium of information drawn from numerous Greek and Arab sources and the medieval geographic writings of John of Hollywood (Johannes de Sacrobosco) and Roger Bacon. Columbus's well-thumbed copy was printed between 1480 and 1483 and bound with two other, shorter cosmographical works. Together these were annotated with 863 postils, the *Imago* itself containing 475. Many of these glossed d'Ailly's discussion of the west-east

extent of the Eurasian landmass. Columbus was seeking the largest reasonable estimate, which would minimize the extent of ocean waters. Ptolemy had estimated that the globe was covered half with land and half with water. Marinus of Tyre, with whom he debates in the *Geography*, accepted the figure passed down from Aristotle through Pliny the Elder and the philosopher and playwright Seneca: three-eighths water and five-eighths land. Somewhat simplistically, Marinus used this to claim that Eurasia extended around five-eighths of the globe, or 225 degrees, leaving 135 degrees of ocean. Columbus started with Marinus's figure and calculated in several other factors to reduce the 135 degrees.

First he added 28 degrees to the mainland to account for the discoveries reported by Venetian Marco Polo in his *Travels*. Polo's famous travelogue through Asia was first published between 1302 and 1314 by Friar Pipino of Bolgna, and Columbus's copy was printed in Antwerp in 1485–1486, after he had presented his plan to João. This contained some 366 notes, most of which were in Columbus's hand. The fame of Polo's book and its centrality to Columbus's enterprise virtually guarantee he had read it before acquiring the Antwerp edition. In any case, Polo told the West about Cipangu, Japan, which he claimed lay 1,500 miles off China's coast, or 26.5 degrees by Columbus's reckoning. He rounded it up to 30 degrees and added an additional nine degrees to account for the Canaries, from which he would launch his voyage across the Atlantic. Doing the math (9 + 28 + 30 = 67) resulted in fewer degrees of open ocean his voyage would cover, or a total of 68 degrees (135–67). Since Marinus claimed that the world was 22,500 miles in circumference rather than 20,400 (Columbus's figure), Columbus cut roughly 10 percent of the number of degrees to an even 60, or one-sixth of the 360-degree circumference.

He found further satisfaction with this figure in the apocryphal Fourth Book of Esdras, a nonbiblical Hebrew religious text given some authority by Saint Augustine, the greatest among early Christian theologians. In 4 Esdras 42 and 47 the author claimed that during Creation, God covered one-seventh of the Earth with water and six-sevenths with dry land. Some of this water, of course, filled the Black, Baltic, and Mediterranean Seas and the Indian Ocean, leaving still less in the Atlantic basin. Columbus's 60-degree distance meant a voyage of 3,400 miles (60×56.66) at the equator and somewhat less as one moved the path north or south. Esdras's quasi-biblical one-seventh meant 51 degrees, or 2,914 miles across the Atlantic, but this would include the distance from Asia to Japan and Europe to the Canaries, reducing open ocean to 51–39 degrees, or a scant 12 degrees or 680 miles, an admittedly absurd figure, even for Columbus. Instead he subtracted the nine degrees (Lisbon-Canaries) from the 51 degrees of Esdras and landed on the round figure of 2,400 Roman miles from the Canaries to Asia, or about 42.36 degrees. He could further justify the reduction by noting that he would not be tracing the equator, along which line the degree is greatest in length, but farther north, where the degree was smaller. If Esdras was correct, Columbus was overestimating the distance.

One last important source that helped Columbus affirm his plan was the famous letter sent by the well-respected Florentine physician and astronomer Paolo dal Pozzo Toscanelli. In 1474 Portugal's King Afonso requested Toscanelli's opinion on the shortest route from Europe to Asia. In response, Toscanelli sent a detailed reply (June 25, 1474) that would have encouraged the king and that ended up in Columbus's hands a half dozen years later. Most scholars reckon that he acquired it through his new noble in-laws who held positions at court in Lisbon. It included a map "that set Columbus's mind ablaze," according to Las Casas. Columbus supposedly wrote to Toscanelli anew around 1480 (Toscanelli died in 1482) and received an encouraging introduction along with essentially a copy of the original letter and a new sea chart that included the "island" of Antillia and Cipangu. The only surviving document, however, is a copy of Toscanelli's letter in Columbus's hand on the flypage of his printed copy of *Historia rerum* by Italian cardinal Aeneas Sylvius Piccolomini (published in 1477 and containing some 877 postils), another of Columbus's cosmographical books. Like Columbus's, Toscanelli's views were shaped by Ptolemy and the ancients, and Marco Polo's influence appears in specific place names. Toscanelli placed "the Indies" directly across the Atlantic from "the Western Isles" (the Canaries). He described the unimaginable wealth of East Asia, and claimed to have actually conversed with an ambassador from China. In the introduction to the disputed second letter, Toscanelli supposedly wrote "the voyage laid down is not only possible but true, certain, honorable, very advantageous, and most glorious among Christians" (quoted in Taviani 1985: 175). The Florentine estimated that the distance to be covered was some 3,000 miles, though interrupted by Antillia and Cipangu, while Columbus stuck by his 2,400-mile calculation.

The importance of the reassurance provided by the writings of Toscanelli, Piccolomini (later Pope Pius II, whose work even discusses the tantalizing prospect of an undiscovered continent), D'Ailly, and others should not be underestimated. Time and again Columbus repeats in his postils "not a great distance to sail" and "short space of sea." Though none of his written sources were composed by mariners, and certainly none had crossed the ocean, several made it clear that such a voyage was not only possible, but would take but a few days. Important cartographers, like Martin Behaim, who in 1492 placed Cipangu in the middle of a rather narrow Atlantic (about 90 degrees, half of which—Portugal to Cipangu—would have been 45 degrees, very close to Columbus's 42 degrees), also provided an encouraging vision of what many others, however, considered impossible.

Assessing Columbus

And so it was that Columbus absorbed the calculations, rumors, estimations, and evidence that supported his quest, and let the rest run off him. He knew both sides of the argument, and importantly there *were* two sides. Columbus did not

convey his conclusions without accumulating the kind of evidence that would sway, if not convince, the royal committees. The only record of Columbus's appearance before King João II, that printed in *Decades of Asia* (1552) by the Portuguese historian João de Barros, describes Columbus as vain, boastful, and a "big talker" whose ideas about reaching Cipangu (or even its existence) were not grounded in fact. Perhaps they were not at that point. Italian historian Paolo Taviani believes that Columbus only consulted the classic texts after having been rejected by João, and traveled to Castile: his initial failure drove him to find confirmatory evidence of the type most likely to sway court intellectuals. De Barros seems to confirm this, as he states that Columbus claimed before João that his plan was based upon Marco Polo's claims, the discoveries made since the days of Prince Henry the Navigator, his discussions with Portuguese mariners about their discoveries, and his knowledge of the experiences of foreign sailors. By the time he pitched his plan to the Castilians in 1486, however, he had found what he had needed. Of course, Ferdinand and Isabella's experts had no agenda except to serve their masters faithfully, and easily accepted the less extreme conclusions of Ptolemy and others that appeared to make a transatlantic voyage unacceptably risky. When the Spanish monarchs finally relented in 1492, it was in the face of the opposition of their own advisers and at the behest of Columbus supporter and chief financial minister Luis Santangel, who counseled that the risk of a few men and ships was well worth the possibility of monopolizing the trade in the wealth of far Asia. Ferdinand had just witnessed the eastern luxury and riches of the defeated Muslims in recently conquered Granada: could he really dismiss the gamble Columbus proposed? And a gamble it was, as the last scene of the Reconquista had all but emptied the royal coffers.

Does this imply anything about unscrupulous collection or massaging of information? Does it in any way imply that Columbus believed one thing and advertised another? Historian Valerie Flint concluded that "it may never be possible accurately to draw a line between that which Christopher Columbus truly believed, and that which he exploited for effect" (Flint 1992: xx). Like the royal experts, Columbus picked and chose his sources and drew the conclusions that suited him best. After all, although Columbus was asking the monarchs for ships and contingent riches, he was placing his own life on the line.

Historically, this was the boldness of Columbus. His was no unique quest: with cartographers, astronomers, merchants, explorers, and kings he shared the vision of reaching Asia by crossing the Atlantic. But he pressed his case with the vigor of a man possessed. His voyages began the first sustained contact between the Eastern and Western Hemispheres since the migration across the Bering ice. Those who decry this contact, who view as an unmitigated disaster the "Age of Discovery" (as far as Europe was concerned, Columbus *did* discover the Americas, whatever he thought them to be), see fit to tar Columbus with every brush they can find. Indeed, recent research has provided strong

evidence that Columbus participated in the Spanish Crown's slave trade with North Africa before his voyage; and many of his actions as expedition leader and colonial administrator were deplorable. Columbus's critics, however, interpret his entire record (for that's all anyone can do) in the dimmest of lights, in our case claiming that he consciously falsified the "cosmographical" evidence to support his agenda. In fact, when claiming that Columbus lied and natives died, are his critics not doing the same kind of manipulation of evidence to support their agenda? This is not to deny or even downplay the importance of the natural and human-made ravages that came in the wake of the Enterprise of the Indies: disease, violence, slavery, dispossession, and cultural destruction are just a few of the tragic and reprehensible elements of the so-called Columbian Exchange. It is, however, to ensure that the historical record is not twisted to serve contemporary agendas, however noble their cause may seem to be.

References and Further Reading

Ayala, Juan de. *A Letter to Ferdinand and Isabella, 1503*. Trans. by Charles E. Nowell. *Commentators*. Trans. and ed. by Lynn Thorndike. Chicago: Chicago University Press, 1949.

Bacon, Roger. *The* Opus Maius *of Roger Bacon*. Trans. by Robert Belle Burke. Philadelphia: University of Pennsylvania Press, 1928; reprinted Sterling, VA: Thoemmes Press, 2000.

Bushman, Claudia L. *America Discovers Columbus: How an Italian Explorer Became an American Hero*. Hanover, NH: University Press of New England, 1992.

Colón, Fernando. *The History of the Life and Deeds of the Admiral Don Christopher Columbus*. Turnhout: Brepols, 2004.

Columbus, Christopher. *The Book of Prophecies Edited by Christopher Columbus*. Trans. by Blair Sullivan. Berkeley: University of California Press, 1997.

Columbus, Christopher. The Diario *of Christopher Columbus's First Voyage to America, 1492–1493*. Edited and trans. by Oliver Dunn and James E. Kelly, Jr. Norman: University of Oklahoma, 1989.

Columbus, Christopher. *A Synoptic Edition of the Log of Columbus's First Voyage*. Trans. by Cynthia Chamberlin and Blair Sullivan. Turnhout: Brepols, 1999.

Davidson, Miles H. *Columbus Then and Now: A Life Reexamined*. Norman: University of Oklahoma Press, 1997.

de las Casas, Bartolomé. *Las Casas on Columbus: Background and the Second and Fourth Voyages*. Edited and trans. by Nigel Griffen. Turnhout: Brepols, 1999.

Fernández-Armesto, Felipe. *Columbus and the Conquest of the Impossible*. London: Phoenix Press, 2000.

Flint, Valerie J. *The Imaginative Landscape of Christopher Columbus*. Princeton, NJ: Princeton University Press, 1992.

Granzotto, Gianni. *Christopher Columbus: The Dream and the Obsession*. New York: Doubleday, 1985.

Koning, Hans. *Columbus, His Enterprise*. London: Latin American Bureau, 1991.

Morison, Samuel Eliot. *Admiral of the Ocean Sea: A Life of Christopher Columbus*. Boston: Little, Brown, 1970.

Nader, Helen, and Rebecca Catz. "Christopher Columbus." In *The Christopher Columbus Encyclopedia* (pp. 161–204). New York: Simon and Schuster, 1992.

Nunn, George E. *The Geographical Conceptions of Columbus: A Critical Consideration of Four Problems*. New York: Octagon, 1977; orig. New York: American Geographical Society, 1924.

Phillips, William D., and Carla Rahn Phillips. *The Worlds of Christopher Columbus*. New York: Cambridge University Press, 1992.

Pliny the Elder. *Natural History*. Cambridge, MA: Harvard University Press, 1944.

Polo, Marco. *The Travels of Marco Polo*. Trans. by William Marsden. New York: Modern Library, 2001.

Ptolemy, Claudius. *Ptolemy's Geography: An Annotated Translation of the Theoretical Chapters*. Edited and trans. by J. Lennart Berggren and Alexander Jones. Princeton, NJ: Princeton University Press, 2000.

Randles, W. G. L. "The Evaluation of Columbus's 'India' Project by Portuguese and Spanish Cosmographers in the Light of the Geographical Evidence of the Period." *Imago Mundi* 42 (1990): 50–64.

Sacrobosco, Johannes de (John of Hollywood). *The* Sphere *of Sacrobosco and Its Commentators*, L. Thorndike, ed, Chicago: University of Chicago Press, 1949;

Sale, Kirkpatrick. *The Conquest of Paradise: Christopher Columbus and the Columbian Legacy*. New York: Knopf, 1990.

Taviani, Paolo Emilio. *Christopher Columbus: The Grand Design*. Trans. by William Weaver. London: Orbis, 1985.

Todorov, Tzvetan. *The Conquest of America*. Trans. by Richard Howard. New York: Harper Row, 1984.

Vignaud. Henry. *Toscanelli and Columbus: The Letter and Chart of Toscanelli*. New York: Ayer, 1971; original pub. London: Sands, 1902.

Wilford, John Noble. *The Mysterious History of Columbus: An Exploration of the Man, the Myth, the Legacy*. New York: Knopf, 1991.

8

European pathogens caused the decline of Cahokia and Mississippian mound builders.

PRO	Chris Howell
CON	James Seelye

PRO

The idea that diseases played a large role in the Columbian exchange, especially in catastrophically reducing pre-Columbian native populations in the Americas, has gained increasing acceptance in the scholarly community beginning with the work of Alfred Crosby in the early 1970s. Crosby popularized the idea of the Columbian exchange as the complex set of exchanges between the Western and Eastern Hemispheres involving people, flora, fauna, technology, and diseases, resulting in an almost complete replacement of American native people in the Western Hemisphere or Americas with immigrants from the Eastern Hemisphere or the Afro-Eurasian landmasses. As the name suggests, most of the research on the catastrophic historical impact of diseases on indigenous peoples of the Americas focuses upon contacts, conquests, and exchanges between the two hemispheres from the time of Christopher Columbus and his four voyages to the Americas between 1492 and 1508.

However, in North America, another group of Europeans, the Norse Vikings who settled and lived in Greenland between 1000–1450, may well have brought an earlier "Norse exchange" that could have contributed to the fall of pre-Columbian North America's most urbanized civilization, the Mississippian mound builders, including the largest known city of native North America, Cahokia, near modern Saint Louis.

Evidence for this intriguing possibility comes from three main sources. First is the Icelandic sagas, oral traditions of Scandinavian voyages and settlements across the North Atlantic, written down between the 12th and 16th centuries as a series of stories, many recorded while the Greenland colony near North America was still in existence. The two most important concerning the Norse in North America are the Greenlanders saga and Erik the Red's saga, written several centuries after the attempted settlement of Vinland in North America by the Norse. The two most important sources for the Norse Greenland colony are Ari Porgilsson's Icelanders Book and the Land Book, both written during the existence of the Greenland Norse settlement between 985 and 1450. As with

165

many oral traditions, numerous written versions were recorded, often leading to varying historical interpretations. However, much like Heinrich Schliemann's rediscovery of legendary Troy using the Homeric epics, Helge and Anne Ingestad used the Icelandic sagas as guides in "rediscovering" the Norse settlement in North America at L'Anse aux Meadows in Newfoundland. The Ingestads used the 16th-century Skaholt map to help predict the latitude of the settlement, indicating just how historically accurate the sagas could be.

The saga evidence led to the finding of Norse settlements in North America, and when coupled with Norse eastern and western settlements in Greenland and Mississippian mound-building sites like Cahokia, represent the second form of evidence, archaeological data. These sites and their associated artifacts are an important crosscheck for the oral and written traditions of both the Norse and the Native Americans as they can be roughly dated and precisely located.

The third form of evidence comes from Native American myths and legends passed down in oral traditions until written down in the later European colonial period. The Wallam Olum or Red Record of the Lenni Lenape or grandfathers of the Algonquin-speaking tribes, including the Delaware Indians, may well be the most important native counterpart of the Norse Icelandic sagas. The Wallam Olum is a record of migrations of Algonquin ancestors into northeast North America, overrunning Cahokia along the way, apparently filling a regional power vacuum created about the time of Norse settlements there. All three forms of evidence—oral traditions, written versions, and archaeological remains—may well point to a much larger impact for the Norse and especially for diseases brought with Norse settlement and trade, into the pre-Columbian world of the Mississippians in eastern North America.

Research into the past is by nature a tenuous proposition with most of the evidence not surviving the ravages of time, and with the rest often frustratingly ambiguous and open to multiple interpretations. Nevertheless, the case will be made here that Norse settlements across the North Atlantic, especially the Greenland settlement established by the family of Erik the Red by 1000, led to repeated contacts and exchanges with Native Americans both via Norse voyages to the North American mainland and via Native American travels and voyages to Greenland. Most of the contact took place between 985–1450, involving the Norse Greenlanders and at least five Native American communities, including the subpolar Dorsett, Inuit, and Thule, as well as Native American tribes in northeastern North America including the Beotuk, Micmac, and Iroquois. Some of these cross-cultural contacts were purposeful, some accidental. All of the contacts had the potential to create "disease vectors" by placing Native Americans, Norse, and disease-carrying domesticated animals from the Old World together, including cattle, pigs, sheep, goats, and dogs. Some exchanges were peaceful, others violent. But all had the potential to spread through close contact the epidemics and pandemics that devastated later Native American populations in the

Columbian exchange. Further, the highly mobile nature of the Inuit, using sleds and dogs by land and skin boats by sea, gave them a range from Alaska to Greenland and from Hudson Bay to the Great Lakes. Since the more mobile groups like the Inuit had only arrived in the Americas from Siberia and Eurasia, they had already been exposed to domesticated animal and Old World diseases not experienced by many Native American groups in North America whose ancestors had been in the region since the last Ice Age, well before the exposure to Old World domesticated animals and diseases. Thus the more immune subpolar groups, often referred to as Eskimos (Inuit, Thule), could pick up diseases associated with the Norse colony in Greenland and spread these diseases by chance along travel routes from Greenland to as far away as the Great Lakes. Finally, the Norse attempts to settle and harvest resources on the North American mainland led mainly by the four children of Erik the Red, Thorvald, Leif, Freyda, and Thorstein, resulted in mobile Norse ships placing cattle and possibly other domesticates on the mainland of North America as early as 1000. Though these settlements ultimately failed, they led to repeated accidental and purposeful voyages to North America for resources such as timber and iron ore during the entire five-century occupation of Greenland by the Norse. There was more than enough contact in both quality and quantity to greatly increase the potential for diseases to migrate from the Old World to the New World, just as happened in the later Columbian exchange (up to 90 percent mortality rates) and possibly with the similar devastating results for Native American populations, of which the largest and most densely populated in North America was at Cahokia.

This interplay of contact, trade, disease, and conflict between hemispheres was most common in eastern North America between 1000–1450. Those time and space parameters fit roughly with the rise and fall of Mississippian civilization, a series of mound-building settlements along the Mississippi, Missouri, and Ohio rivers' drainages. These sites had the most urbanized, least mobile populations of natives in pre-Columbian North America, and thus were the most likely to be devastated by any disease episodes brought into the region along land or water trade networks. Possibly associated with the Muskegon language family of natives, these mound builders also had a lineage that placed them in eastern North America as far back as 3,000 years ago, meaning most had not been exposed to Old World disease pandemics and epidemics as the later arriving Inuit were. Most of the mound-building sites associated with the Adena, Hopewell, and Mississippian cultures were linked by river canoe travel on the Mississippi, Ohio, and Missouri rivers' drainages and included were part of a complex trade network linking densely urbanized sites spanning eastern North America and including the Great Lakes and Saint Lawrence River. This meant the Mississippian civilization arose around the time of the attempted settlement of Vinland by Leif Erikson and other children of Erik the Red and would later decline just as the Norse Greenland settlement declined.

All of the circumstantial evidence suggests just how likely it must have been that these Norse settlers probably unknowingly brought early disease episodes into North America. If evidence can place diseased Norse and their farm animals in close proximity to Native North Americans between 1000–1450, the probability of pandemics or epidemics increases dramatically.

Further, if those disease vectors are likely and common among the Norse Greenland settlements, they would be transferred to Native American populations such as the Dorsett, Inuit, and Micmac and Beotuk whom the Norse came in contact with in the subarctic. The Dorsett and Beotuk territory declined upon arrival of the Norse, but the hardy and mobile, more disease-resistant Inuit and Micmac expanded into it using sled dogs and skin boats, which represented an ideal means by which disease might be spread far from Norse population in rapid fashion. Both the subpolar Inuit and the more temperate climate Micmac represented ideal vectors by which pathogens could travel and emerge into the Mississippian world south of the Great Lakes. Both groups migrated from Siberia across the subarctic well after Old World pandemics and epidemics associated with domesticated animals and urban populations had repeatedly ravaged the Old World. The Inuit in particular would have been much more immune to disease the Norse brought in from the Old World than would those at older lineage cultures such as the mound-building communities. The Inuit and Micmac would thus be the likely carriers as would their sled dogs and possibly the wild caribou herds in terms of location between the Norse in Greenland and the mound builders at sites like Cahokia. It would be the Old World pathogens carried on Inuit and Micmac dogsleds and skin boats into the vast trading network of rivers and lakes controlled by the Mississippian mound builders where we might expect the biggest disease impact.

The devastating impacts of later Columbian exchange disease episodes such as the 1540s Hernando de Soto expedition into the southeast of North America are well documented and historically look similar to earlier Norse and native interactions. The difference is the Norse settlements were permanent, longer-lived, and far larger than these later expeditions. Additionally, the potential disease vectors and trade networks between Greenland and the Mississippian homelands had a much greater geographic reach than later European expeditions in the 16th century. The Norse Greenland settlement lasted 450 years, numbered over 5,000 people, and contained several thousand domesticated cattle, sheep, goats, pigs, and dogs—all known carriers of later disease pandemics that would ravage Native American populations in the Columbian exchange. Numerous Norse expeditions were sent to North American lands known as Helluland (probably Baffin Island), Markland (probably Labrador), and Vinland (probably Newfoundland and the Saint Lawrence River areas south) with known voyages between 985 and 1350. Further, the highly mobile Inuit were visiting Greenland between 1000 and 1500, both with dogsleds and by skin boats. Norse voyages to North America from Greenland for timber, iron ore, ivory, and furs along with

Inuit travel to Greenland for Eastern Hemisphere trade goods like metal and glass ensured contact across five centuries.

Why were Old World diseases so devastating to many Native American populations? Research on the role of diseases in the process of reducing Native American populations by up to 90 percent in the pre-Columbian Americas focuses on post-Columbus times, especially 1500–1800. This research has convincingly made the case for repeated pandemic and epidemic disease episodes, usually associated with Eastern Hemisphere domesticated animals such as cattle, sheep, goats, pigs, and chickens in forms such as smallpox, chickenpox, cattle pox, and swine, sheep, and chicken flu, catastrophically reducing Native American populations who had few animals to domesticate (turkeys, dogs, llamas and other camelids, guinea pigs, and a few ducks) and thus almost no built up immunities to diseases that had ravaged the Eastern Hemisphere over 10,000 years of domesticated animal history.

Such long-term, high-population, and high-contact rate disease vectors and their catastrophic impacts on indigenous populations are well understood for later Columbian exchange contacts. The Spanish conquests of Mexico and Peru are historic examples in which permanent settlements from the Eastern Hemisphere, with its 10,000 years of domesticate animal diseases flowing into Native American populations, resulted in multiple pandemics and epidemics across three centuries that reduced some population to a tenth of their numbers. Eventually, the question of a "Norse exchange" with Native North America may not be one of the role of disease, but one of just how large that impact was, if other historical examples may serve as a guide.

Evidence from Norse Sagas

The Scandinavian peoples, mainly from modern Norway, Sweden, Denmark, and Finland, began to expand out of their homeland in waves possibly associated with climate and population changes during the European Middle Ages from the 7th to the 14th centuries. Though often associated with Viking raiding and trading activities, these migrations and invasions were often about finding new land for settlement, whether it was Swedish Rus and Finns in Russia and even Byzantium; Danes in England, France, and the Mediterranean; or Norwegians in Ireland, Iceland, Greenland, and, eventually, North America. Most expeditions were actually planned by the Northmen or Norse, who traveled by ships along coasts looking for raiding, trading, and settlement opportunities. The Norse reached the British Isles by the 8th century, including Ireland, and then moved into the Faeroes Islands and Iceland by the ninth century. It was mainly Norwegians and some Celtic populations who settled in these areas, bringing domesticated animals like sheep, goat, pigs, cattle, and Norwegian elk hounds along with a host of Eastern Hemisphere diseases.

From this historical context emerged Erik the Red, a powerful but outlawed clan leader, who was being forced out of Iceland due to poor behavior, even for a Viking! It would be his family who would eventually settle Greenland and attempt to settle and explore parts of North America. Our evidence comes mainly from the Icelandic sagas referred to as the Greenlanders saga and Erik the Red's saga as well as occasional references in other sagas to places such as Greenland, Helluland, Markland, and Vinland and Hop. In reality, these sagas were originally oral traditions passed down for several centuries before they were put in written form, often associated with the literary movement in Iceland.

Erik sought to escape his dilemma in the Old World by exploring to the west and creating a new settlement in lands to the northwest of Iceland. Previous explorers like Gunnbjorn, son of Ulf the Crow, were consistently blown off course from Iceland to the west, in the early ninth century. Even a relative of Erik's family, Naebjorn Galti attempted to explore the area around 978, though he was murdered in a mutiny. Erik left Iceland and found this land (southern Greenland), which he explored for three years before returning to Iceland to recruit for a settlement expedition. Erik the Red named the land Greenland to help attract settlers to this area that faced North America across the Labrador Sea. His expedition left for Greenland with about 25 boats, of which about 14 safely arrived just after 984, indicating just how dangerous North Atlantic waters were. Some no doubt were blown off course, possibly toward North America, a common problem for other Norse navigators in these waters. In the next decade three more fleets followed to Greenland, and by 1000, over 500 settlers along with domesticated animals arrived to set up a larger eastern settlement and a smaller western settlement of perhaps 100 people. They found an empty land that had clear signs of previous occupation, probably by the Dorsett people of sub-Arctic North America. Greenland was in the midst of a climatic warming trend, and grazing land was plentiful for cattle, pigs, sheep, and goats until a cooling trend emerged in the 14th century.

According to Erik the Red's saga, following settlement in Greenland, Erik the Red's son Leif Erikson left for Norway to serve the king and report on Greenland. The king sent Leif back to Greenland to spread Christianity, but on Leif's return voyage to Greenland his crew was blown off course, landing somewhere in North America. He described the area as full of wild wheat, berries, and grapevines, with large maple trees, naming it Vinland the Good according to Erik the Red's saga. Leif also rescued a lost ship of other Norse, indicating just how common it was to be blown off course to North America when heading for Greenland or Iceland in the North Atlantic.

However, the Greenlander saga gives credit to Bjarni Herjólfson for sighting North America. Bjarni came seeking his father in Greenland only to be blown off course to North America. Supposedly, Bjarni did not let his crew on shore until he sailed north and found Greenland. It is then that Leif bought Bjarni's

**L'Anse aux Meadows: Europeans in America
Nearly 500 Years before Columbus**

Very little doubt remains that the Norse sailors arrived in the Americas centuries before Christopher Columbus and even built permanent structures. At the very northern tip of the island of Newfoundland is L'Anse aux Meadows, which is now preserved by Parks Canada and recognized as a United Nations World Heritage Site. At L'Anse, the remains of several timber-and-sod buildings attest to the Viking presence in the region. Around the year 1000, Leif Erikson, son of the Norse King Erik the Red, landed at the site, naming it Vinland. Others followed Leif, constructing the permanent structures found there today. One commonality that they shared with later European explorers was conflict with the indigenous populations of the region in which they settled, which forced them to leave the area and go back to Greenland within a decade.

ship and made an exploratory voyage to Vinland, returning with news of pasture, berries, and days with more sun than in Greenland or Iceland.

Erik the Red's other son, Thorvald, then took Leif's ship and attempted to settle the land Leif had described. He found Vinland and Leif's camp called Leifsbodir (possibly L'Anse aux Meadows) where he and his crew of 30 explored for several seasons. A fight broke out with natives and Thorvald died of an arrow wound and was buried in Vinland. Contact with the Native Americans, possibly Beotuk near Labrador, represented an early opportunity for disease spread, but it would not be the only time. Thorvald's brother Thorstein Erikson, sought the body for burial, but was blown off course and forced to return to Greenland. Thorstein died upon return to Greenland as part of an epidemic fever outbreak, one of many Norse Greenland disease episodes documented in the sagas and indicating just how virulent Greenland would have been as a disease-breeding area.

Thorstein's wife, Gutrid, apparently survived the time of fever, and then wed an Icelandic merchant named Thorfinn Karlsefni. According to Erik the Red's saga it was Thorfinn who led a large settlement expedition to Vinland consisting of between 60 and 250 people on at least two ships. Even Erik the Red's half daughter, Freydis, who had married one of the expedition members, went along. The group sailed north along Greenland and then west to Baffin Island (Helluland) before turning south to the North America mainland of Labrador (perhaps Markland). Here they let ashore two Celtic runners to explore the land to the south where the land was declared good, and the expedition wintered, releasing cattle ashore but having to rely upon whale meat when starvation set in. Eventually spring brought better fishing and hunting, and the expedition, still seeking Vinland, split into two, with a small group heading north and the main group under Thorfinn heading south. The northern group was blown off course to

Ireland where most perished. The south group fared better, settling near a lake that flowed down into a river and out to sea. They named it Hop and built a settlement there. Thorfinn's son Snorri was born then, indicating the presence of infants, and possibly infant disease susceptibilities to minor respiratory infections that could easily be picked up by natives from the Norse. This would further increase the breadth of diseases natives would have been exposed to.

Natives were encountered at Hop. These were possibly the Micmac who inhabited the region from Labrador to the Great Lakes and Saint Lawrence River and who have legends of blue-eyed people bringing misfortune. These natives in nine canoes left for the south. That winter, once again, cattle were allowed to pasture freely, increasing the likelihood of encounter with the Native Americans who were by this time watching the foreign Norse at Hop. In spring the Native American canoes came in large numbers, bringing hides for trade and seeking red cloth and weapons. Thorfinn Karlsefni forbade trade in metal weapons with the Skraelings as the Norse called them, but did offer milk, perhaps cow's milk, which the Native Americans consumed. This encounter and the consumption of milk, a completely new drink for the natives, presented a prime opportunity for disease spread, either human to human or animal to human, as the small Norse cattle, including a bull, still roamed freely. The bull scared off the natives who came back en masse, and a battle broke out between the two groups. The natives even brought and used simple catapults, driving the Norse into rocky crags. The presence of catapults in the battle probably indicates just how far south the Norse were, perhaps even in the Saint Lawrence River area. Catapults were generally used on wooden stockade villages like those associated with the Algonquian and Iroquois speakers or even with Mississippian sites. They were not associated with warfare among small nomadic bands of Inuit in the subarctic area. One wonders if some disease was not already setting in among the native people, and they resolved to drive out the foreign Norse who brought such troubles to their land. Many Skraeling and several Norse were recorded as killed in the battle. The Norse expedition thought it wise to move back north, killing five more natives along the way before settling at a land they called Straumsfjord. Thorfinn Karlsefni took the boat and searched for the northern expedition without success. Thorfinn then returned to Hop for several months, but Freydis and others remained behind, possibly constructing a second settlement, which Thorfinn returned to that winter. All told, the expedition stayed three winters before leaving Vinland. On the way back to Greenland they captured a family of natives and brought them back to Greenland. Apparently two native boys did survive the diseases, probably indicating an increased level of immunity as suggested for the Inuit. The natives did mention that larger tribes with kings lived inland from them, possibly a reference to Great Lakes area tribes. Thorfinn Karlsefni eventually returned to Iceland with his son who was born in Vinland, Snorri. Snorri's offspring included the bishop of Iceland, and it is in Iceland at

that time the sagas concerning Greenland and North America become oral traditions. Vinland or Winland was mentioned in the work of Adam of Bremen in 1075 and indicates that information about the area was being disseminated back to Europe even as the Icelandic sagas were still in oral forms.

Finally, Erik's daughter Freydis attempted to settle Vinland yet again with partners Helgi and Finnbogi. However, after apparently heading back to Vinland settlements already constructed by Leif, Freydis conspired to kill the rival settlements of Helgi and Finnbogi and their people, personally doing away with the five women. Freydis and her group failed to stay and returned again to Greenland where Leif found out about the killings but took no action.

Eventually Erik the Red's children and relations led five separate voyages to the south and west, labeling the territories Helluland, Markland, and Vinland in the sagas. One settlement they established at L'Anse aux Meadows in Newfoundland has been excavated and could be the Leifsbudir or even Hop settlement of the sagas, if indeed those were separate settlements. It eventually had eight buildings and up to 80 settlers along with farm animals. A variety of Norse artifacts were excavated at this site, including 99 iron nails and cape brooches.

Reflecting on the Norse voyages to North America between 970 and 1030, several pieces of evidence suggest the likelihood of disease spread and possible implications for the rise of Mississippian civilization at about the same time. First, several hundred Norse encountered at least an equal number of natives, whether Inuit in the north, or Beotuk and Micmac in the south, over the equivalent of several years. The contact was intensive, including trade, warfare, close human and animal contact, and consumption of foreign food and drink. The red cloth traded by Karlsefni's expedition proved ideal in the later Columbian exchange for transport of diseases like smallpox. Certainly, the consumption of milk by natives and contact with the cattle released to graze provided ample opportunity for disease to transfer from animal to human. The taking of natives back to Greenland certainly resulted in exposure where a number of disease episodes such as the fever that killed Erik's son Thorvald had already impacted the Norse. Norse saga references seem to place them in the area northeast of emerging Mississippian civilization, probably around the Saint Lawrence River mouth. The references by the natives taken back to Greenland indicate denser populations into the interior as well. If a major outbreak hit Native Americans at this time, it might have led to population decline or territorial changes that allowed expanding Mississippian civilization to grow into these more north and easterly regions, or at least gave them less impeded room for growth along major waterways. Because of differential mortuary ritual practices and the catastrophic nature of diseases, population estimates are really no more than "guesstimates" for much of Native America at this time. Later records from the Columbian exchange indicate that fields of Native American remains might be found after diseases entered their communities, with no one to perform the burial rituals that archaeologists so rely upon for population
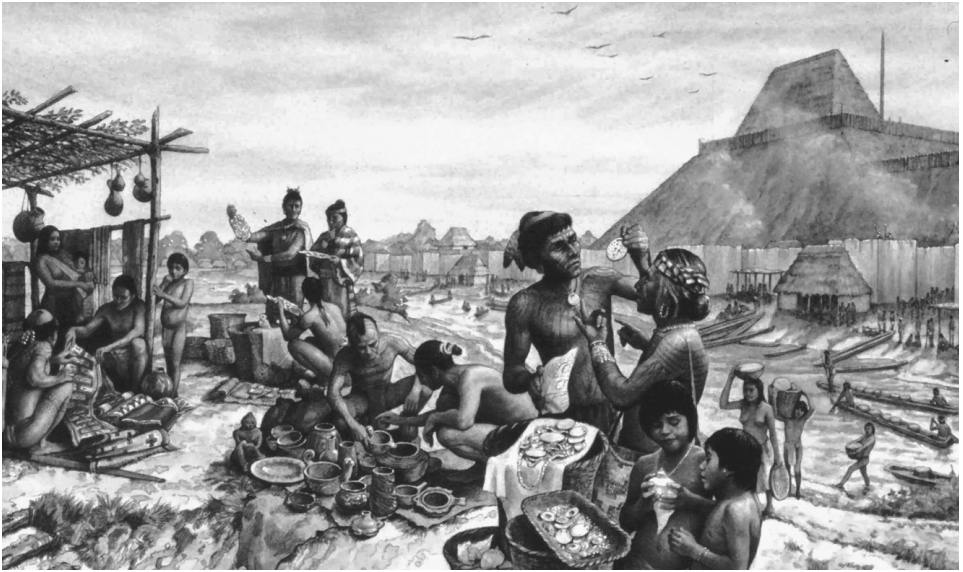
Artist's rendering of a market at Cahokia, about 1150–1200 CE. Cahokia was the largest prehistoric city north of Mexico and was a religious, political, economic, and cultural center. (Cahokia Mounds State Historic Site, painting by Michael Hampshire)

preservation and estimation. Population estimates for pre-Columbian North America range anywhere from less than 1 million to over 12 million, and this after a century of research and debate. So population estimates are of little use when compared to changes in territorial power for this era.

Population estimates aside, all of the early exploration and settlement did leave tangible clues of Norse and Native American interactions across eastern North America and at least hint at the possibility of correlation between the rise of Mississippian civilization and the entrance of the Norse. A silver penny made in Norway before 1080 was found during excavations at an archaeology site in Maine where it had been modified into a pendant. The site and others inland also contain special forms of chert or stone for making tools, found nowhere else but in Labrador. Such long-distance trade networks extended inland to the Great Lakes and Ohio River, providing easy paths by which disease vectors may have emerged.

Contact directly into North America by diseasecarrying Old World people and animals by 1000 makes this a prime date to look for territorial changes among Native Americans in eastern North America due to disease outbreaks. On the eastern seaboard, the rise of Inuit and Micmac power bases, replacing the shrinking territory and power of older Dorsett and Beotuk peoples, fits well with their proximity to Greenland Norse colonies. The rise of Mississippian civilization from older mound-building civilizations of the Adena and Hopewell on the Ohio River drainage also fits chronologically. Even Cahokia, the largest site in pre-Columbian

North America, is a close fit. Its meteoric rise and control of the river system from the confluence of the Missouri and Mississippi rivers near modern Saint Louis began in the 10th century. Cahokia sits many miles north and upstream from the mouth of the Ohio and Mississippi rivers' confluence and thus was not in direct line with any disease epidemics brought in from the attempted Norse settlement of Vinland and North America. However, their territorial reach would soon grow to include areas such as the Ohio River, placing Mississippian civilization in the direct path by which diseases from the Norse settlers might spread by water below the Great Lakes region.

Voyages to Greenland, mainly from Iceland, would take place for the next four centuries, with the last recorded voyage taking place around 1407. The Norse Greenlanders were dependent on Norway for most ships and iron, because so few iron ores and timber were available in Greenland. This meant voyages when possible to Markland for those resources and more contact with Native Americans. Greenland exported walrus and narwhal ivory, seal and walrus skin rope, polar bear pelts, as well as falcons to Europe. Thus disease vectors in both directions were present during the entire period of the Greenland colony. However, cooling climates and increasing ice lessened the number of voyages by as early as the13th century. Worse yet, in Greenland itself the sagas and archaeology finds indicate the Norse response was to bring in more food supply such as their favorite meat, pigs, along with cattle for dairy products, horses for transport, and sheep and goats for marginal grazing lands. All these animals proved lethal as disease carriers and vectors for later Native Americans encountering Eastern Hemisphere people in the later Columbian exchange. Pigs, for instance, apparently were devastating to natives on the de Soto expedition of 1539–1543. Thus Mississippian civilization probably arose just as the most likely time for a disease epidemic or even pandemic was receding due to the failure of the Vinland Norse colony by Eric the Red's children. But a disease incubator of sorts, the Norse Greenland colony itself, with its 5,000 members and even more domesticated animals, represented a constant threat to the north, a threat that became worse as the Norse colony fell upon hard climatic times and more desperately sought resources from both North America and from ships back in Europe.

Some Norse arrived sick in Greenland, and disease episodes in the colony were common throughout its existence. A larger eastern settlement and a smaller western settlement were established, both in the southwest area of Greenland, where richer grasslands allowed for more farm animal feed and longer growing seasons. As the climate cooled, the Norse food supply was reversed, with terrestrial foods increasingly replaced by maritime foods. Although few fish bones are found in Norse Greenlander trash middens, the Norse usually made fish "porridge" out of the fish, eating bones and all, especially in drought times. These trends, too, brought the ever more malnourished and sickly Norse and their animals into increasing contact and conflict with the Inuit over the maritime food

supply, providing further opportunity for disease to spread. Such hard times were ideal for the spread of the Black Death and bubonic plagues that devastated Europe and even Iceland by 1402 and beyond. The status of the Black Death, whether it hit Greenland, is still unknown, but other pandemics and epidemics from later Columbian exchange times might provide a model for what the Black Death could have done to Native American communities.

The Norse did not have to visit North America in order to spread disease. Inuit legends and records by Ivar Bardarson, the steward of church property in Greenland, indicate the Inuit helped knock out the last of the Norse in the western settlement, some 300 miles northwest of the larger eastern settlement around Erik the Red's estate at Bratthalid. Bardarson, in the *Story of the Greenlanders*, relates that he took a force to the west settlement in the mid-1300s to push back the Inuit, who were apparently moving down from north Greenland as the climate worsened. Competition over the northern hunting grounds for ivory was at the center of the conflict. He found only a deserted settlement with the cattle, sheep, goats, and horses roaming wild. Later, both Inuit legends and sagas indicate even the eastern settlement was raided by the Inuit, with 18 Norse killed. Such close contact did not require Norse voyages to North America for diseases to spread. The Inuit themselves could become the primary vector.

Interestingly the major port of the eastern settlement was Herjolfsnes, closest to the harvest of ivory in the north, closest to the vital resources of timber and iron ore in Markland, and also the most used port of the Greenland colony for European ships. Even after Scandinavian ships stopped calling at the eastern settlements, other European ships may have continued to arrive at the port based on clothing styles discovered at the local church cemetery. Basque traditions hold that they established a whaling port in Newfoundland around 1372. Perhaps English and German pirates visited Greenland as well. They raided Iceland and may have traveled to Greenland just before 1400, about the time the bubonic plague hit Iceland hard.

An Inuit legend recorded by Niels Egede, a Dane who grew up in Greenland, supports this late interaction by European raiders into the Greenland Norse settlement. The legend indicates three ships from the southwest came to plunder and killed some of the Norse who in turn managed to capture one ship before the other two escaped. The next year a larger pirate fleet arrived and plundered extensively. Norse survivors sailed south, leaving some behind. In the third year yet another pirate fleet arrived off Greenland, and the remaining Inuit and Norse who could fled to the highlands, returning in the autumn to find that farms had been destroyed and the last inhabitants taken away. If such a legend has some historical basis, it opens up even more opportunities for disease vectors into Native American communities. First, it indicates Inuit and Norse may have been living or trading side by side. Second, pirate ships may have continued on into North America or been blown off course there when returning to Europe. Third,

the Greenlander Norse may have used the captured ship to launch yet another voyage for timber to Markland. All are possibilities that increase the number of potential contacts for disease spread from European to Native American populations. As late as the 1540s, ships bound for Iceland were being blown off course to Greenland. One crew found what may be the last surviving member of the old Greenland Norse colony. He was an old man found dead with his knife and patched clothing at the edge of the sea.

The last historical references to Greenland begin with a familiar story—a merchant ship to Iceland blown off course to Greenland. This leads to a Christian ceremony for the wedding of Sigrid Bjornsdottir and Thorstein Olafsson, one of the merchants on the ship, at the Hvalsey Church or main church of the eastern settlement in 1408. The merchants also witnessed a burning at the stake of an adulterous man. The church was tied to the papacy in Rome and had become quite powerful in the Greenlander affairs since its inception with Leif Erikson in the 11th century. Some missionary activity to Vinland was also noted but has never been confirmed via archaeology. Pope Alexander VI suggested a voyage to reestablish contact with the Greenlanders, whom he believed had not seen a ship for some 80 years by the late 14th century. However, other records indicate ships did occasionally arrive and that Norse Greenlanders did occasionally leave. Nicolus Germanus records the Danish cartographer Claudius Clavus as visiting and mapping Greenland in 1420 (the Danish Crown had technically taken over Greenland in the late 1300s). Peter Groenlindiger records a Norse Greenlander as visiting Norway in 1426. It is entirely possible that on some of these ships Europeans and others who were not Norse at all were visiting and were blown off course to Vinland or other parts of North America. This opens the possibility of even further disease transmission opportunities without diseases needing to originate or pass through an existing Norse Greenland colony.

One story from the sagas indicates just how common this may have been. This is the legend of Great Ireland, a land supposedly located many days' sail west of Ireland and near Vinland the Good. According to Ari the Wise of Iceland, his ancestor Are Marson had been visited in Great Ireland by ships blown off course around 983 and further that a man named Bjorn Asbrandson who set sail from Iceland around 1000 was also blown off course and resided in Great Ireland. Neither could escape and instead were employed with great esteem by the local chieftains or Indians. Gudleif Gudlaugson, according to the Eyrbyggja saga, was shipwrecked on Great Ireland in 1029, and his party was rescued and then sent away by Asbrandson who advised them not to return as the locals were apt to change their minds often about strangers on ships. While these stories have yet to be confirmed archaeologically, they indicate just how common contact between natives and Europeans could be, regardless of the Greenland Norse colony.

Regardless of when the last Norse ceased to exist on Greenland, the dates are of great interest concerning the demise of Mississippian culture. It is in the

mid-15th century that Algonquin and Iroquois legends indicate a time of troubles and separation from ancestral homelands. Archaeology data from Mississippian sites generally indicate that many Mississippian sites in the north begin to decline or were abandoned during this same century. It is to the Native American myths and legends as well as archaeological evidence that we now turn to for confirmation of the impact of the Norse Exchange in mainland North America.

Native American Evidence

The Dorsett culture associated with Baffin Island, Greenland, and parts of subarctic Canada, were the native inhabitants when the Norse Greenland settlement was founded by Erik the Red just before 1000. Interestingly they had already abandoned Greenland, according to the ruins found by Erik's settlers. Soon they would lose control of much of northeast arctic North America due to the combined presence of the Norse to the east and the Inuit from the west. Possessing neither the sleds and hide boats of the Inuit nor the longboats of the Norse, Dorsett people may well have been reduced by both European pathogens from sick Norse and their domesticated animals brought en masse along with Norse settlement. The power vacuum was filled by the incoming Inuit, ideally suited to the subarctic environment, with adaptations ranging from igloo housing and sled dogs to bows and arrows, and custom skin clothing and hide boats with waterproof suits.

The Inuit Eskimos first replaced the retreating Dorsett culture by around 1000 and then, eventually, after repeated interactions, came to dominate the declining Norse settlements in Greenland by the 16th century. Thus the Inuit, with their superior land and sea mobility, territory, and trade routes stretching from Norse Greenland down to Newfoundland and the Great Lakes, were an ideal disease vector or conduit for Old World, Eastern Hemisphere diseases. This was true when the Norse first arrived just as Mississippian settlements began to boom, and it was true during the Norse decline in Greenland, just as Mississippian settlements were in decline. The Inuit were the ones who killed off the last of Norse in the western settlement, according to both Norse and Inuit sources. They were in close quarters with both disease-ridden Norse and their sickly animals. Inuit traders and trade goods could travel by land down to the Great Lakes region and the Beotuk and Micmac territories, and by sea even farther.

Inuit stories were gathered by anthropologist H. J. Rink in 1858, as illustrated by Inuit artist Aron of Kangeg. The Inuit called the Norse the "Kablunet" and had both cooperative and competitive relations with them. Inuit stories make clear that they were living and working with the last of the Norse Greenlanders, perhaps in the larger, eastern settlement. Because the Inuit had spread quickly from their Siberia origins across subarctic North America, immunities had developed in their ancestry to domesticated animals back in Eurasia. Thus they were probable disease carriers but not necessarily decimated by Eurasian diseases from Norse

contact. However, the groups south of the Great Lakes, especially the widespread traders of the Mississippian civilization that emerged from the much older mound-building tradition, had no such immunities.

The Micmac controlled much of Quebec, Labrador, and Newfoundland at the time of the Norse Greenland settlement. Like the Inuit in the north, who expanded with the arrival of the Norse Greenland, the Micmac expanded into northeastern North America just as Erik the Red's children attempted to settle Newfoundland and Vinland. The Newfoundland Beotuk Indians declined on Newfoundland just as the Dorsett had in Greenland. Such patterns of power changes in regions are difficult to tie archaeologically directly to disease, but they are suggestive. Certainly we can say that as soon as the Norse arrived, major shifts in population and territorial control began to take place associated with the Norse in Greenland and the Norse in Vinland. The territory impacted stretched from the Saint Lawrence River and the Great Lakes in the south to the Labrador Sea in the northeast, directly across from the Norse Greenland western and eastern settlements.

Further evidence of just how significant the power changes were in eastern North America comes from the Lenni Lenape or the ancestral Delaware Native Americans. Their legends indicate they migrated in an easterly migration route from Siberia into northwest North American and then down through Mississippian territory, possibly Cahokia, finally arriving east of the Great Lakes and into the Iroquois area south of the Huron and Micmac tribes. Their legends are one of the few to have survived the Columbian exchange intact and in permanent symbolic form on wooden tablets with red symbols known as the Wallam Olum. Considered the "grandfathers" of most Algonquian-speaking tribes, the most widespread of all linguistic families of Native Americans in North America, their oral traditions are unique for their breadth and scope. Today the last of the Lenni Lenape are in Bartlesville, Oklahoma, due to the reservation system movements of the historic period. Researcher David McCutchen, who worked for 15 years on the translation of the Wallam Olum, translated the story based partly on the wood tablet symbols and partly on oral stories confirmed by the elders of the tribe in 1980.

If the Wallum Olam or red record is accurate, it may contain evidence of the Lenni Lenape journey into the plains of North America, where their small nomadic populations allied with early Iroquois to push out the last of the Mississippians at Cahokia in a joint migration and attack. The record makes clear that a large city was conquered near the Missouri and Mississippi rivers' confluence, and that its descendants eventually fled south, possibly to reemerge as the mound-building Natchez of historic times. Natchez legends indicate they were founded by people moving from the north along the Mississippi after a great catastrophe. Could this have been the one described in the Wallam Olum? According to that record, the Lenni Lenape moved through the lands of this great city with permission, but the city king became alarmed at how many people were moving through and attacked them. The Lenni Lenape allied with other enemies of the city, the

Iroquois to the east, and eventually defeated the urbanites, forcing them to move south.

Several strands of the legend are worth looking at in detail. First, if this site was Cahokia or another large Mississippian site that was head of a Mississippian chiefdom or territory, their population would normally be far greater than a no-madic, migrating group like the Lenni Lenape or Delaware could overcome. Second, only a weak and declining Mississippian civilization would have allowed such migrating groups through its territory or would have assessed the migrating tribal population to be large versus their sedentary farming population. Third, the Lenni Lenape were migrating into the northeast for a reason, perhaps because of a power vacuum or population decline in the mound builders. Finally, since they were nomadic and similar to the Inuit in their late arrival from Asian (Siberian) origins, the Lenni Lenape, like the Inuit, were not as vulnerable to old diseases as the older mound-builder populations might have been. All of this may indicate a much-reduced Mississippian population, reeling from the effects of population reduction, possibly due to disease episodes. Archaeological data indicate many mound-building sites (Aztalan, Toltec Mounds, Marksville, Cahokia) were well fortified and dealt with war on a regular basis. It seems unlikely that migrating groups would have the capability of knocking out the most prolific native civilizations in North America unless something was amiss internally. Interestingly, after the Delaware moved into the Mississippian lands, they began to use many of the mounds and geoglyphs in their sacred ceremonies even though they were originally nomadic and without any known architecture on the plains. William Romain asserts that the Delaware eventually utilized mound-builder structures for yearly renewal ceremonies as if the Delaware had taken over the old mound-building areas. Again this seems highly unlikely that a newly arrived migrating group could just push aside a two-millennium mound-building tradition and appropriate their most sacred sites unless something else such as a catastrophe had taken place to weaken the Mississippian mound builders.

Although much debate by archaeologists still exists as to exactly who and what qualifies as Mississippian, the term is used here to describe most of the cultures of the Mississippi and Ohio rivers' drainages as well as those in southeast North America who were associated with the later prehistoric mound sites. Their tradition stems from earlier mound builders known as the Adena and Hopewell along the Ohio River basin. Sites such as Aztalan in Wisconsin, Cahokia in Missouri, Etowah in Georgia, and Moundsville in Alabama are all classic examples. Because these densely populated sites are often connected by major river valleys as well as integrated land trade networks, disease vectors were more likely via these transportation routes. Certainly, at Cahokia, the large numbers of elite burials associated with the early occupation of the site are not noted for the later occupation period, when stockade walls and water control problems become the norm. But no matter how well or poorly researchers using archaeological data

can place the rise, fall, and territory of native cultures in North America, certain problems cannot be overcome, at least at this point in time.

Conclusion

The biggest issue in dealing with ancestral North American communities is that the populations were always somewhat mobile or semisedentary. This makes it extremely difficult to determine actual population as Charles Mann (2005) has summarized in the nearly century-long attempt by the research community to grasp demographic data for pre-Columbian North America and indeed the entire Western Hemisphere before 500 years ago. Thus implications of population decline or impact as diseases possibly moved in via Norse, Norse animals, and native contact are difficult to prove or disprove. So while we can make inferences about the Mississippians and about general changes in power distribution across native North America, with the arrival and settlement of the Greenland Norse colony we cannot be sure that they are tied together. However, the timing is extremely intriguing. The arrival of seafarers from the Old World or Eastern Hemisphere in the 10th century fits well with the rise of Mississippian civilization, especially its largest set of earthworks at Cahokia near modern Saint Louis. Remember, many of the Norse ships arrived accidentally after being blown off course, so we cannot be sure when the possibility for first contact and early disease vectors came into existence. Iceland, for instance, had been settled since the eighth century, and the Norse sagas indicate just how common it was for seafarers heading to and from Iceland to find themselves blown off course to Greenland or North America instead. Without question, the attempts to settle Vinland by Erik the Red's children placed all the necessary elements for disease spread from European to Native American communities directly in North America during the 10th century.

Further, the decline of Mississippian civilization may well be tied to the decline of Norse Greenland. In the 14th century the Norse Greenland western settlement declined and was overrun by the Inuit. This historic scenario places highly mobile and partly immune Native Americans directly inside of increasingly sickly and diseased Norse communities that are on the decline. The same holds true for the final 15th-century collapse of the larger eastern settlement, including Erik the Red's old estate at Bratthalid and the main church at Hvarsley. Either one of these episodes or the increasing ship traffic from Europe culminating in the Columbus expeditions of 1492–1508 could have produced the right scenario to spread diseases into the Native American Mississippian world, helping to bring about its downfall. The most likely places to have been hit would have been the places with the most contact points and the largest concentrated populations, an apt description for Cahokia, which sat with its river canoe trade fleets at the conjunction of the Mississippi and Missouri rivers, just upstream from the Ohio River. The Ohio River flowed from the lands of northeastern North America,

where Norse diseases would have first hit native communities that had no immunities, unlike the Inuit communities who had some immunities due to relatively recent travel from Siberia, and due to their 450-year proximity to Norse Greenland. We must not forget too, the Black Death and bubonic plague outbreaks that swept into Europe in 1347 near Naples, Italy. By 1402 even Iceland was devastated, and it seems possible that Greenland too, or even North America may have been hit, especially if diseased sailors or their clothing, a common conduit for the spread of the plague, made it to resource-poor Norse Greenland or to inquisitive communities of Native Americans via ships blown off course from Iceland or seeking resources from Vinland or Markland for Greenland.

While it cannot be conclusively demonstrated on the level that modern civilization can track modern pandemics and epidemics, the likelihood that major disease episodes, as part of a "Norse exchange," probably hit native North American populations is strong based on a number of factors. First, the impact of later disease episodes often associated with sick immigrants and farm animals from the Eastern Hemisphere is now well documented and demonstrated for the later Columbian exchange as the work of Noble David Cook (1998) and Alfred Crosby (1972) has demonstrated. Similar historical circumstances existed for the Norse exchange. Second, the presence of sporadic Norse–Native American interactions for almost half a millennium greatly increases the opportunities for disease dissemination. Over 5,000 Norse and several thousand farm animals were eventually present in Greenland and some even briefly in North America at Vinland from the 10th to the 13th centuries. Third, the timing of the Norse in Greenland and the rise and fall of Mississippian civilization is remarkably coincidental. The lack of historical writing for Native American communities means we can only tighten down the timelines so much. But they do fit together at present. Fourth, more immune and mobile groups such as the Inuit and Micmac provided hosts and disease vectors whereby diseases originating in the Old World traveled through the Norse Greenland colony into subarctic native populations and finally into the densely populated Mississippian realm, where little immunity existed. Finally, both Norse and Native American lore, history, and archaeology indicate the near constant presence of disease in some form in Norse Greenland. Occasional contact, whether accidental or purposeful, by land or by sea, provided the opportunity for disease vectors to become major shapers in parts of the pre-Columbian world. What remains now is for definitive evidence, probably archaeological in nature, to provide one positive example, much as finding L'Anse aux Meadows in Newfoundland did to legitimize the stories of Vinland in the Icelandic sagas.

The major roadblock is perhaps just how much has been lost in historic times concerning the Columbian exchange and concerning Mississippian and mound-building civilizations. These sites would be primary in looking for the initial paths of epidemics into Cahokia and Mississippian worlds in general. Just

think if the written records and archaeological evidence no longer existed for the initial introduction of the Black Death into Europe via Mongols besieging Kaffa in the Black Sea, or for the Mediterranean merchants carrying it into the south Italian peninsula. We might also be debating if the Mongols had anything to do with the mysterious changes in European power distribution and population about the time of the Renaissance as we now debate about the Norse role in pre-Columbian North America.

References and Further Reading

Cook, David Noble. *Born to Die: Disease and New World Conquest, 1492–1650*. Cambridge: Cambridge University Press, 1998.

Crosby Jr., Alfred W. *The Columbian Exchange: Biological and Cultural Consequences of 1492*. Westport, CT: Greenwood, 1972.

De la Vega, Garcilaso. *The Florida of the Inca: The De Soto Story*. Translated by John and Jeannette Varner. Austin: University of Texas Press, 1988.

Diamond, Jerod. *Collapse: How Societies Choose to Fail or Succeed*. New York: Penguin, 2005.

Fowler, Melvin. "Cahokia Atlas." Illinois Historical Society, Springfield. 1989. Available at http://libsysdigi.library.uiuc.edu/oca/Books2008-06/cahokiaatlashist00fowl/ (accessed June 10, 2010).

Jones, Gwyn. *A History of the Vikings*. Oxford: Oxford University Press, 1984.

Mann, Charles. *1491: New Revelations of the Americas before Columbus*. New York: Alfred A. Knopf, 2005.

McCutchen, David, trans. *The Red Record: The Wallam Olum*. Garden City Park, NY: Avery, 1993.

Pauketat, Timothy. *Ancient Cahokia and the Mississippians*. Cambridge: Cambridge University Press, 2004.

Roesdahl, Else. *The Vikings*, rev. 2nd ed. Translated by S. Margeson and K. Williams. New York: Penguin, 1998.

Romain, William. *Mysteries of the Hopewell: Astronomers, Geometers, and Magicians of the Eastern Woodlands*. Akron, OH: University of Akron Press, 2008.

Sephton, J., trans. "The Saga of Erik the Red." Translated from the old Norse 'Eiríks saga rauða' in 1880. Available at http://www.sagadb.org/eiriks_saga_rauda.en (accessed June 10, 2010).

Waskey, Andrew. "Micmacs." In *The Encyclopedia of North American Colonial Conflicts to 1775*, Vol. 2. Edited by S. Tucker. Santa Barbara, CA: ABC-CLIO, 2008: 484.

CON

Cahokia was an impressive settlement. It flourished for centuries before the arrival of Europeans. It was settled in approximately the 10th century and was finally abandoned around 300 years later. Scholars have long studied and disagreed over the causes of the decline and abandonment. Reasons that have been postulated include warfare, disease, internal and external rebellion, and environmental factors. There is also a contingent that feels Cahokia only declined because of the introduction of European pathogens to which the residents had no natural immunity. While it is true that between 90 and 95 percent of all indigenous peoples died within 100 years of first contact with Europeans, the evidence is clear that Cahokia declined and was in fact abandoned long before its inhabitants could have ever breathed in the toxins Europeans exhaled.

To get the best picture of why Cahokia declined and was ultimately abandoned, the reader is best served by looking to ethnohistory. Ethnohistory is a blend of disciplines, particularly history, anthropology, and archaeology. Taken individually, no single discipline can provide a compelling enough explanation for the question asked here. However, taken together, the evidence is overwhelming. Scholars and researchers have proven the reasons Cahokia was abandoned, and it was a sum of factors, all centered on environmental degradation.

Cahokia was located on the middle Mississippi River floodplain, a good location as far as the environment was concerned. It was located near a diversity of smaller habitats that included open water and other wetlands, which provided a source of protein in fish. However, more important, the soil in the region was optimally suited for corn cultivation. Along the floodplain, Cahokia grew and residents grew multicrop gardens (Yerkes 2005). This strategy worked for a long time and may have continued to work if Cahokia had remained a small, relatively dispersed area. However, as it continued to grow and thrive, it became a religious and cultural center for the Mississippian peoples. Its population rose, perhaps to as many as 30,000 inhabitants at its peak, and the environment could not handle it. Demands on wood and fuel for construction increased to the point the environment suffered irreparable damage. Erosion and runoff increased, and flooding worsened, particularly during the summer growing season. The consequences of this, including economic and social, proved disastrous for Cahokia, especially when the crops failed. As a result, the inhabitants of Cahokia were left with no choice but to disperse to other areas.

Cahokia was located in an area known as the American Bottom. The American Bottom is a floodplain located just below the confluence of the Illinois, Missouri, and Mississippi rivers, and has a variety of woodland landscapes, including some that are submerged part of the year. There are many lowland areas that are constantly under the threat of flooding. There are also many higher, well-drained areas and prairie land. As a result the inhabitants of

Cahokia had a variety of plant and animal life to help sustain them, in addition to the rich soil.

Evidence uncovered by researchers indicates that the inhabitants of Cahokia had a fairly diverse diet of fish, amphibians, reptiles, mussels, birds, and mammals. Their findings point toward fish being a primary food source, with deer being a distant second. They may have eaten dogs on occasion as well. The residents also gathered a variety of nuts, seeds, fruits, and greens. They also cultivated tobacco, squash, and gourds. Interestingly, beans have not been found, so it seems the people of Cahokia only partook in two of the three sister crops. Corn was the number one food source, and it was the most widely produced and consumed (Yerkes 2005). In the years leading up to the emergence of Mississippian groups, hunting became more efficient with better weaponry, namely the bow and arrow.

Corn was far from new to the region Cahokia developed. Evidence shows that it first appeared in the area at least 1,000 years before Cahokia emerged. However, it appears that corn took hundreds of years to catch on and become a staple of the local diet. Once people developed more static communities, the populations in individual areas exploded. A reliable, stable food source was needed, and corn provided well. Altogether, the diversity in dietary options led the residents of Cahokia to enjoy large annual feasts. They served many purposes, including many political functions, tribute, alliance building, and warfare. In addition, the evidence show that feasts at Cahokia included specialized goods made by skilled artisans at the request of the elites and the ruling chiefs.

A number of settlements were located in the region around Cahokia. They were all part of what anthropologists refer to as the Mississippian peoples. They were well known for building elaborate mounds. Cahokia, as the main religious and political center, had the largest number of mounds, and there were clearly delineated areas for royalty, ceremonial use, and class boundaries (Ambrose et al. 2003). Cahokia eventually became the paramount location, and its rivals either joined with Cahokia or were removed through warfare. At its peak, Cahokia was the largest population center north of the Rio Grande prior to European contact.

Monk's Mound

The focal point of Cahokia, both in its prime and today, is Monk's Mound. Named after a community of Trappist monks who lived in the vicinity and grew vegetables on the first terrace of the two that remain, its 16-acre base is larger than any of the Egyptian pyramids. At the top of the mound, a wooden temple that was over 100 feet long was home to the 'Great Sun,' the spiritual and temporal ruler of the city. Between Monk's Mound and the nearby Mound of the Ruler-Priest, a plaza of almost 50 acres formed the central ceremonial location for most of the over 10,000 Cahokians, not to mention their main sports arena, where *chunkey*, a two-player game played with a stone wheel and spears, was played.

Monks Mound at Cahokia in southern Illinois near St. Louis, Missouri, is the largest pre-historic earthen structure in the Americas. It is 100 feet high and covers more than 14 acres and contains an estimated 22 million cubic feet of earth. Monks Mound was at the center of the Cahokia community, which included about 120 mounds over a five to six square-mile area. (Cahokia Mounds State Historic Site)

The mounds at Cahokia and of the Mississippian people that are so well known were constructed with baskets of earth dug from areas that seemed to exist for that purpose. Some of those dirt sources have been located today. It seems apparent that there was an overall predetermined plan in place in order to manage such a large project that had a great number of laborers involved. By the time Cahokia was completed, the surrounding landscape and environment had been transformed. The overall size of land covered was impressive, and at least 120 mounds were constructed. Monks Mound is the most impressive mound that survives. It is a large, multileveled platform located within what used to be the primary ceremonial area.

Taken together, it seems that the most compelling evidence for the decline and abandonment of Cahokia was a congruence of factors. Overall, the separate factors led to extreme environmental disruption. These factors include changes in climate, warfare, leadership shortcomings and failures, internal strife, and a disruption of long-standing relationships with nearby groups. However, it seems that in the search for the one main cause of Cahokia's abandonment, the issue of food takes center stage. Within the region of Cahokia, the procurement and

production of food was focused around the location itself. This was possible because of the variety and diversity of habitats within the floodplain and other nearby areas. This strategy of maintaining sustenance works well for dispersed populations. However, take the same environment as Cahokia was located and factor in the concentration of large numbers of people, and the consequences were devastating over time (Young 2000).

Anthropological evidence provides a mass of evidence that within Cahokia and its surrounding area, widespread environmental degradation took place and exerted an overwhelming impact on the capability of the area to produce food for its residents. This was exacerbated by myriad other factors discussed previously. For example, demands on wood and fuel resources were intense. After the construction of Monk's Mound, a massive palisade was built around the central portion of Cahokia, and excavations have shown it was then rebuilt at least three times. A number of additional palisades have been discovered at other sites near Cahokia, although on a smaller scale (Dalan 2003). These, too, show evidence of being rebuilt. Furthermore, a number of burnt dwellings were located. This provides ample evidence for both internal and external strife, all factors in the decline of Cahokia, not to mention the increased stress on the environment.

Archaeologists and anthropologists have concluded that the residents of Cahokia used a great deal of wood. Mississippian people built structures known as wall-trench buildings, and each of those used an average of 80 posts. If a household of five people had two structures each, a population of 25,000 residents would have used at least 800,000 wall posts. A tree certainly would have provided more than one post, but they were also used for roofing, furniture, watercraft, palisade construction, and fuel. Finally, untold numbers of trees were cut down to make room for crop production. All of these factors impacted the environment greatly, especially when it appears that the region did not have all that much wood to begin with. The terrain of the American Bottom is a mixture of habitats. These include lakes, ponds, creeks, floodplains, and prairies. Prairies may have occupied half the land. Trees were largely restricted due to topographical reasons. A variety of trees grew in the region, including oak, hickory, cottonwood, willow, elm, ash, sycamore, mulberry, and walnut. Experts estimated the total number of trees in the region at approximately 600,000. The best woods in the region for building and fuel would have been rather rapidly used up in the earlier stages of Cahokia's development, leaving those who came later to make due with what was left.

Water issues hastened the decline of Cahokia as well. The overexploitation of wood created many issues the inhabitants could not have been aware of. Due to its location, Cahokia would have experienced serious flooding after each heavy rain. Water runoff would have increased, and the land's capacity to absorb excess water would have decreased. In addition, the flooding would have been devastating to the cornfields that the inhabitants planted in the bottomland area around the settlement. Cahokia relied quite heavily on the production of

that crop, and its reduction and eventual ruin would have been catastrophic for its residents.

The agricultural capacity of Cahokia was destroyed and literally smothered during the later years of its occupation. The fields surrounding it were sloping, and the higher woodlands in the area had been disturbed too rapidly for natural replacement. With every rainstorm came more heavily eroded surfaces. One side effect of that was the deposit of less and less fertile soil in the cornfields. Furthermore, Cahokia Creek and its smaller tributaries were changed into ever widening, yet shallower, channels. The sum of these factors is a higher risk of flooding during the summer, just when the corn crop was growing at its peak. The higher levels of water and the increased amount of poor sediments infiltrating the soil would have resulted in not only flooding, but also large pools of standing water. The soil became saturated, and the crops would have literally suffocated. Finally, the residents would not have been able to rely on game such as deer, because the local population would have felt the effects of environmental degradation as well, and their numbers would have been eradicated.

The social and economic factors of these issues were staggering and ultimately insurmountable. The main issue was the decline in crop production and its final failures. Cahokia was in the perfect position to experience the worst effects. It was located downstream and had the largest concentration of residents in the entire area. All of those residents depended heavily upon a reliable harvest. However, they could not and did not foresee the consequences of their actions upon the environment, and as such could not and did not take any preventative measures. Once the severe rate of environmental degradation began, it was too late to stop. Scholars and researchers have uncovered ample evidence to prove both the rapid deterioration in the conditions of both life and the environment around Cahokia, and the fact that the residents of the area reacted to the reality of their situation by abandoning it and starting anew elsewhere.

Archaeologists and anthropologists also discovered other events that took place simultaneously with those described above. It appears that many residents of the region decided to move to higher ground within the area, and there was also a general trend of migration back into the interior upland areas. By approximately 1350, the complex community centered at Cahokia had been abandoned. This is long before any of its inhabitants were exposed to European pathogens, as has been postulated by some scholars.

There was no one single cause of Cahokia's abandonment. Rather, it was a confluence of causes, all described above. The three main factors were the size and concentration of Cahokia's population, the availability of wood in the region, and the effects of erosion on agricultural capabilities, particularly with regards to corn. Furthermore, the abandonment of Cahokia did not occur overnight. It appears it took between 50 to 100 years, during which time people gradually left. The crisis of Cahokia began during what scholars refer to as the Stirling phase

(1100–1200). It gradually became worse, and reached a climax during the Moorehead phase (1200–1275). By the end of the Sand Prairie phase (1275–1350), Cahokia was nearly completely abandoned.

References and Further Reading

Ambrose, Stanley H., Jane Buikstra, and Harold W. Krueger. "Status and Gender Differences in Diet at Mound 72, Cahokia, Revealed by Isotopic Analysis of Bone." *Journal of Anthropological Archaeology* 22 (2003): 217–26.

Chappell, Sally A. Kitt. *Cahokia: Mirror of the Cosmos*. Chicago: University of Chicago Press, 2002.

Dalan, Rinita A., et al. *Envisioning Cahokia: A Landscape Perspective*. DeKalb: Northern Illinois University Press, 2003.

Woods, William I. "Population Nucleation, Intensive Agriculture, and Environmental Degradation: The Cahokia Example." *Agriculture and Human Values* 21 (2004): 255–61.

Yerkes, Richard W. "Bone Chemistry, Body Parts, and Growth Marks: Evaluating Ohio Hopewell and Cahokia Mississippian Seasonality, Subsistence, Ritual, and Feasting." *American Antiquity* 70, 2 (2005): 241–65.

Young, Biloine Whiting, and Melvin L. Fowler. *Cahokia: The Great Native American Metropolis*. Champaign: University of Illinois Press, 2000.

9

Shakespeare's plays were written by someone other than William Shakespeare of Stratford-upon-Avon.

PRO   Alexander Hugo Schulenburg
CON   Jeffrey Mifflin

PRO

William Shakespeare is universally considered the greatest playwright in the English language and his works are performed and studied the world over, both in their original language and in translation. Based on documentary evidence from both his lifetime and the decade after his death, the author William Shakespeare is commonly identified with William Shakespeare of Stratford-upon-Avon, England, a local trader and property investor who was also involved as a shareholder and actor in the London theater company that performed most of Shakespeare's plays.

However, doubts about the authorship of Shakespeare's works were first raised in the late 18th century by the Rev. James Wilmot, rector of Barton-on-the-Heath, to the north of Stratford, who was puzzled by the lack of local references in his works. Since the mid-19th century, those doubts have grown steadily and have resulted in a veritable library of books on the Shakespeare authorship question. Doubters have even included notable actors like John Gielgud, Derek Jacobi, Jeremy Irons, and Michael York. Among the academic establishment, however, the Shakespeare authorship debate is almost entirely frowned upon. Others ask what difference a knowledge of Shakespeare's true identity makes to his works or to one's appreciation of them.

The authorship question falls into two separate though consequential parts: Are there reasonable grounds for doubting that William Shakespeare of Stratford was the author of Shakespeare's plays?, and if so, Who is more likely to have been their true author? John Michell's (1996) *Who Wrote Shakespeare?* is highly recommended as a balanced, single volume introduction to the debate.

A brief word on terminology is in order. The claim that William Shakespeare of Stratford-upon-Avon was the author of the plays published under his name is usually referred to as the "orthodox" position and its proponents as Stratfordians. Those who question the orthodox position have been referred to pejoratively as heretics, deviationists, and anti-Stratfordians, but usually refer to themselves with reference to the alternative author they propose, such as the Baconians (who propose Francis Bacon), the Oxfordians (who propose the Earl of Oxford), and the

191

Marlovians (who propose Christopher Marlowe). There are also the Groupists who propose that the works were a collaborative effort. In order to distinguish William Shakespeare of Stratford from the author of the works, anti-Stratfordians sometimes spell his name Shakspere, the spelling he used himself, while referring to the author as Shake-speare, a hyphenated version found on some editions of his work. For ease of reading, this chapter will distinguish between William Shakespeare the author and Shakespeare of Stratford the man. References to 'Shakespeare's works' are always to be read as to the author.

Shakespeare of Stratford's Biography

A William Shakespeare (at that time spelled Shakspere) was baptized on April 26, 1564, in Stratford-upon-Avon, England. He was the son of a local glove maker, who in later years, held the office of bailiff of Stratford. William Shakespeare's adult life was divided between Stratford, where he had a wife and children and was a local tradesman and property investor, and London, where he had a share in a theater company, the Lord Chamberlain's Men, for whom he also acted, and where he had property interests in two theaters, the Globe and the Blackfriars. Shakespeare always maintained business interest in Stratford, where he died on April 23, 1616. He is buried in Stratford's Holy Trinity Church. A monument near his reputed burial place shows him holding a quill and paper.

The documentary evidence pertaining to Shakespeare of Stratford's family, property interests, and involvement in court cases is reasonably extensive, whereas evidence of his life in the London theater is rather meager. Explicit contemporary evidence that Shakespeare of Stratford was a playwright is nonexistent.

Listing all 12 known references to his career in the theater may go some way to explaining why there is indeed an issue over whether the Shakespeare of Stratford, who was a man of the theater, was also William Shakespeare the playwright:

March 15, 1595: Shakespeare is paid, alongside William Kempe and Richard Burbage, fellow actors in the Lord Chamberlain's Men, for performing two comedies before the queen.

1598: Named as one of the principal "comedians" acting in Ben Jonsons's *Every Man in His Humour*.

1599: Named as one of four lessees of the site of the Globe Theatre, and named elsewhere as being in occupation, along with others, of the Globe.

1602: Named in a diary entry in connection with his fellow actor Richard Burbage.

1603: Named as one of the principal "tragedians" acting in Ben Jonson's *Sejanus*, and also named in a royal permit authorizing him and his associates to perform plays.

1604: Named in accounts as one of several actors who received scarlet cloth to provide a uniform for the royal procession through London.

1608: Named with Burbage in a document pertaining to the transfer of the Globe Theatre.

1613: Paid, together with Burbage, for the design and manufacture of a shield to be carried in a tournament.

1615: Named in a court case as a shareholder in the Globe and in property in Blackfriars.

1616: In his Last Will and Testament Shakespeare of Stratford makes small bequests to John Heminge, Richard Burbage, and Henry Condell, all of the King's (formerly the Lord Chamberlain's) Men.

While this evidence leaves no doubt that Shakespeare of Stratford had been involved in the London theater, none of these contemporary references connect him with playwriting. Although a number of plays and poems were published under the name of William Shakespeare during that time, and although a number of contemporary writers referred to a poet and playwright by the name of William Shakespeare, there is no direct evidence linking the two. That is, there is no evidence from Shakespeare of Stratford's lifetime that says that he was also the writer William Shakespeare.

Instead, the principal evidence linking the two is the posthumous collection of Shakespeare's plays published in London in 1623 (seven years after Shakespeare of Stratford's death) and titled *Mr. William Shakespeares Comedies, Histories, and Tragedies*. This volume, commonly called the First Folio, contained 36 plays, some of which had been published previously, either anonymously or under Shakespeare's name, but some of which are published therein for the first time (Shakespeare 1623). The volume was apparently collated by John Heminge and Henry Condell, two of Shakespeare of Stratford's former theater colleagues, who also wrote a preface for it. Moreover, in one of the several other commendatory poems to the volume, the author Shakespeare is referred to by the playwright Ben Jonson as "sweet swan of Avon" (that is, of the river Avon, that runs through Stratford-upon-Avon), while another, by Leonard Digges, speaks of Shakespeare's "Stratford monument" (presumably referring to his church memorial).

Given the above, what reasons could there possibly be for questioning the identification of William Shakespeare the playwright with Shakespeare of Stratford? Arguably the main reason for doubting that Shakespeare of Stratford was the author of Shakespeare's work is that the life and character of Shakespeare of Stratford, as they appear from the documentary record, do not match the life and character that most expect the author of Shakespeare's works to have had. That is, Shakespeare of Stratford does not appear to have the background, education, life experience, or outlook that would most likely have enabled and inspired him to

write the works of William Shakespeare. After all, as John Southworth has pointed out, "Shakespeare was able to mould whatever material came his way to an aesthetic expression of his own unique experience of life and of the world around him, or to do so in words that at their finest and best reach to universal truths" (2000: 7).

Anti-Stratfordians do not necessarily claim that someone of Shakespeare's humble social and educational background could not have written the plays. Indeed, they propose someone of his background might well have acquired the necessary knowledge, either formally or informally, as well as the life experience on which his works draw, but there is no evidence that Shakespeare of Stratford ever did. Despite this, for some anti-Stratfordians only a member of the aristocracy could match the characteristics they expect of Shakespeare, a claim that Stratfordians condemn as outright snobbery.

Shakespeare of Stratford evidently did not shine at whichever school he may have attended, and as it is known for certain that he did not attend university, he cannot, as a pupil, have been singled out for the kind of patronage that enabled others of his background to attend university, such as the playwright Christopher Marlowe, his exact contemporary, who went to Cambridge University, but who had come from a very similar family background, being the son of a shoemaker.

Instead, what can be gleaned about Shakespeare of Stratford from the documentary record of his activities in Stratford and in London provides a consistent picture of a small-town businessman with material and social ambitions, who showed little interest in education. He does not appear to have been known locally as an author, and there is no record of his family ever having shown any interest in or awareness of his supposed literary fame (no copy of the First Folio, at the very least, is known to have been owned by them; in fact, no books belonging to Shakespeare of Stratford have ever been found).

There are also striking contrasts in the apparently rather conventional outlook of Shakespeare of Stratford and the outlook reflected in Shakespeare's plays. In *Henry VI, Part II* (Act 4, Scene 7), for example, the rebel Jack Cade accuses Lord Say of having "corrupted the youth of the realm in erecting a grammar-school," to which Lord Say replies:

> [. . .] ignorance is the curse of God
> Knowledge the wing wherewith we fly to heaven. (Craig 1891)

Shakespeare of Stratford, despite his wealth, is certainly not guilty of Cade's charge, as he made no educational bequests, either during his lifetime or in his will, and at least one of his daughters was illiterate. By contrast, Edward Alleyn, the first great actor of the Elizabethan era, went as far as founding Dulwich College in London, which thrives to this day.

Given that the available material on the life of Shakespeare of Stratford is insufficient for full-length biographies, his biographers have resorted to writing primarily about the time and world of Shakespeare (both the man from Stratford and the author), rather than about the man himself, although the plays and poems are often drawn on to speculate about changes in Shakespeare's opinions and emotional life. Yet, there is nothing in the known life of Shakespeare of Stratford to explain the generally accepted developmental pattern of his work, such as his final period with its emphasis on redemption, reconciliation, and hope. As Brenda James and William Rubinstein have stressed, there is a "complete lack of any nexus between [Shakespeare of Stratford's] life and the evolution and development of the plays" (2005: 27).

Last but not least, the undocumented "lost years" of Shakespeare of Stratford between 1585 (when his twins were born) and 1595 (when he is listed as a member of the Lord Chamberlain's Men) allow his biographers room to speculate on how those years may have been spent, usually with a view to explaining how the man from Stratford acquired the prodigious knowledge and experience displayed in his work. These lost years have thus variously been claimed to have been spent as a law clerk, a private tutor, a soldier, and so forth, or even several or all of the foregoing. Nor is absence of evidence treated as indicating that something did *not* happen, but it is instead relied upon for stating that something *may have* happened, as there is nothing to say that it didn't. In the absence of circumstantial evidence in favor of these assertions, such speculations are, however, wholly unwarranted and bad historiography.

More crucially, assuming that Shakespeare of Stratford did indeed acquire the relevant knowledge and life experience, there is nothing about his documented activities (excepting the plays and poems, if they are his), which would indicate that he was the man who had grown from being the son of a small town tradesmen into the dramatist whose extraordinary mind and imagination one can encounter in his plays.

One particularly strange attempt by Stratfordians to dismiss such arguments about the lack of Shakespeare of Stratford's learning and life experience is in effect to denigrate Shakespeare's output, claiming that in his own time Shakespeare was viewed merely "as a writer for the common class" (Matus 1994: 290). As for the mismatch between Shakespeare of Stratford's life and his supposed writings, it has been argued correctly that Shakespeare's plays and poems not only drew on existing works, but that his writing lay firmly within the literary and theatrical conventions of his time, thus the history plays and the vogue for Italian settings. Shakespeare therefore emerges as little more than a brilliant craftsman, who wrote near formulaic plays as a means of making a living. This is in stark contrast with traditional scholarship on Christopher Marlowe, who is generally portrayed as a great innovator in the field of English drama. Indeed, Stratfordians have discussed at length the debt that Shakespeare's output owes to Marlowe.

Shakespeare the Poet and Playwright

Despite what Stratfordians may claim, no contemporary references to Shakespeare of Stratford identify him definitively as a playwright. Although a "William Shakespeare" is referred to as a playwright and poet, such references are of little value in themselves, as they establish no connection with Shakespeare of Stratford. For example, Francis Meres writes in his *Palladis Tamia* in 1598 that "As Plautus and Seneca are accounted the best for Comedy and Tragedy among the Latines: so Shakespeare among the English is the most excellent in both kinds for the stage" This, however, only shows that Meres knew plays to have been written under the name William Shakespeare, whoever may be hiding behind that name. The same applies to entries in an account book of the Master of the Revels, which, alongside some entries of performances at court of several Shakespeare plays, list a Shaxberd in a column headed "The poets which mayd the plaies." In turn, records of payments to Shakespeare and his fellow actors for the performance of plays reveal nothing about their authorship.

Only one reference is usually held up as showing definitively that Shakespeare of Stratford the actor was also a playwright. In his *Groatsworth of Wit*, published in 1592, the playwright Robert Greene refers to an actor *and* playwright as "Shake-scene," which has been taken by most orthodox scholars as a reference to Shakespeare of Stratford. It is also supposedly the first time Shakespeare is alluded to in the context of the London theater and literary scene (surprisingly, Shakespeare of Stratford, who most orthodox scholars argue began his career in acting, does not appear on any lists of actors prior to 1598). Instead, there are good grounds for arguing that the reference is to the actor Edward Alleyn, rather than to Shakespeare of Stratford.

Notwithstanding the popularity and unrivaled quality of Shakespeare's plays, neither Shakespeare of Stratford nor anyone by the name of William Shakespeare is known to have moved in the courtly and literary circles, such as Raleigh's, in which his works and evident intellectual qualities would have been

Known as the 'Cobbe portrait,' this painting by an unknown artist is believed to be the only portrait of William Shakespeare painted during his lifetime. The painting was unveiled in London in March 2009. (AP Images)

appreciated. Instead, Shakespeare of Stratford appears to have moved only within the circle of his fellow actors and among the people of his hometown. Was this merely because Shakespeare of Stratford was low born and had no university education, although he eventually attained the status of a gentleman, or because he did not in fact possess any of the intellectual qualities evident in the plays and poems of Shakespeare?

Moreover, the death of Shakespeare of Stratford passed unnoticed by the literary and theatrical establishment, and the only bequests in his will to anyone associated with the world of the theater were to three of his fellow actors: John Heminge, Richard Burbage, and Henry Condell. Most notably, there is no bequest to Ben Jonson, the playwright who speaks so highly of his friend Shakespeare in the First Folio.

Given the doubts about Shakespeare of Stratford, a large number of individuals have over time been proposed as the true authors of Shakespeare's works. The following sections will examine in brief the kind of evidence that writers have been looking for in support of their respective authorship contenders and will present examples of the evidence for some of them.

Searching for Evidence

As weak as the case for Shakespeare of Stratford may be, there is of course no direct evidence in support of one or another of the alternative authorship candidates, such as a signed manuscript, although some have argued that hidden ciphers in Shakespeare's works reveal their true author. Those claims have received particular ridicule by Stratfordians, mostly deservedly so. In their search for evidence, Oxfordians, Marlovians, and others have hence had to focus on circumstantial evidence in support of their chosen candidate. Such evidence is usually sought in the autobiographical aspects of Shakespeare's works, in their display of personal knowledge and experience, as well as in similarities in the style and themes of Shakespeare's works and that of a particular authorship candidate.

Ciphers and Cryptic Clues

Supporters of several candidates, particularly of Francis Bacon, have placed particular emphasis on the presence of ciphers or cryptic clues in Shakespeare's work, which they claim reveal the true author of those works. These are also said to be found in the works of some of Shakespeare's contemporaries, who apparently knew of the authorship secret.

Oxfordians have, for example, suggested that the cryptic dedication of the 1609 edition of *Troilus and Cressida*, "A never writer, to an ever reader," is a play on Oxford's name, E. Vere (Edward de Vere). They have argued the same with reference to the word "every" in the line from Sonnet 76 that "every word doth almost tell my name."

Hidden Messages

The crypto-analysis techniques used by the anti-Stratfordians have been used over the centuries to attempt to extract secret messages from the Bible, among other texts. As early as the 13th century, rabbis and Jewish mystics were using such techniques to uncover revelations in the Old Testament, and more recently, Michael Drosnin's best-selling *Bible Code* attracted a great deal of sensational attention.

Skeptics point out that we privilege the random occurrences we find meaningful, and so would attach more import to the word 'bacon' being 'discovered' in a Shakespearean text than the word 'pork.' To demonstrate the essential meaninglessness of such findings, Australian television host John Safran used a similar process to point out 'hidden messages' about the September 11 attacks in the lyrics of Vanilla Ice.

As Elizabethans were fond of ciphers and word puzzles, however, the suggestion that ciphers and other cryptic information may exist in Shakespeare's works cannot be dismissed out of hand.

Autobiography

Whereas most Stratfordians have over time come to deny that any of Shakespeare's works are autobiographical, not least because of the insurmountable mismatch between those works and the documented life of Shakespeare of Stratford, anti-Stratfordians have placed considerable emphasis on what they believe to be the autobiographical aspects of Shakespeare's writings. Oxfordians have arguably done the most work in this field, scouring the plays and poems for matches with events in their candidate's life. Oxford is hence noted to have lost £3,000 to a London merchant named Michael Lok (or Lock), just as Antonio, in *The Merchant of Venice*, posts bond for 3,000 ducats with Shylock. When Jacques is told in *As You Like It* (Act 4, Scene 1), that "You have sold your own lands to see other men's," this is just what Oxford once did to finance his extended stay on the continent.

The problem with autobiography as evidence of authorship, however, is that such evidence is only convincing if the autobiographical references or allusions pertain to information that would only have been available to the author. If someone else had access to that same information, perhaps because it was the subject of court or tavern gossip, one may simply be looking at biography, not autobiography. Even as biography, such information is still of interest, however, as it limits the field of authorship candidates to those who moved in the circles in which the relevant biographical information could have been obtained. Would Shakespeare of Stratford, for instance, have heard of the Earl of Oxford's marital troubles? Might it have been the subject of general gossip or would he have had to move in aristocratic circles?

Another area of autobiographical evidence is the question of what Shakespeare's works tell us about the author's outlook on life, society, politics, and so forth. This has been discussed by Stratfordians as much as by anti-Stratfordians, and there is little agreement on the matter. Was the author of Shakespeare's works a secret Catholic or not, was he bisexual, did he look down on the common class, and so forth? In the absence of any agreement on these issues, the answers to these questions can easily be made to fit the known views of any of the alternative authorship candidates. Oxford is thus said to match Shakespeare's aristocratic outlook, while Marlowe's supposed homosexuality is said to match the sexual orientation of the author of the *Sonnets*.

Personal Experience

Although Shakespeare's works may or may not feature autobiographical references or allusions, they are usually said to be based at least in part on personal experience, whether of hunting, seamanship, the law, or life at court, although there has been extensive debate on whether all such knowledge may instead have been derived from hearsay or books.

Arguably, the principal area of investigation has been whether the author of Shakespeare's works had traveled extensively both in the British Isles and on the continent, or whether information on the settings of his play, whether in Scotland, Denmark, or Italy, was derived solely from travelers' tales and books. That research has mostly been independent of the merits of any particular authorship candidate, and even Stratfordians have suggested that Shakespeare of Stratford may have traveled to Italy during his "lost years."

There is indeed ample evidence of an acquaintance with Italy beyond mere book learning or secondhand accounts. For example, in *Othello* (Act 1, Scene 1) Shakespeare writes of "special officers of night," showing knowledge that in Venice there was a special night police called "Signori di notte," while in *The Merchant of Venice* (Act 3, Scene 4) Shakespeare employs the local Italian term *traghetto*, when he writes "Unto the traject, to the common ferry / Which trades to Venice." Also, details found in Shakespeare's *Venus and Adonis* are derived from a version of Titian's painting of Venus and Adonis that Shakespeare could only have seen in Italy. Lastly, Shakespeare was clearly influenced by the Italian *comedia dell'arte*, as is evident especially in *Love's Labour's Lost*.

Oxford, for one, is known to have spoken French and Italian and to have traveled extensively on the continent. Although some writers claim apparent geographic errors in Shakespeare's works as evidence that he had never traveled at all, others have shown that such apparent errors can be explained after a little research. Shakespeare consequently did not err when he places a harbor in Verona, as that city was located on an extensive system of inland shipping canals, which were essential to regional transport.

Style

From the perspective of literary rather than historical analysis, the comparison of Shakespeare's style with the works of other writers is one of the most interesting areas of investigation, but also the most difficult for anti-Stratfordians to tackle.

Oxfordians have compared letters and poems in Oxford's own name with the works of Shakespeare and have found notable similarities, while explaining the apparent difference in the quality of the poems by reference to their youthfulness, as opposed to Oxford's mature works as Shakespeare. Oxford's *The Rejected Lover*, for example, has been compared to Shakespeare's *Lucrece*.

Marlovians have arguably made the strongest case out of all the candidates where the question of style is concerned. While the claim that Marlowe eventually wrote under Shakespeare's name is dismissed by those scholars who detect a considerable difference in thematic emphasis between the two writers, it is unintentionally supported by those orthodox scholars of Marlowe who see a thematic progression in the works of what they believe to be by two different writers. A. W. Verity (quoted in Pinksen 2008: 6), for example, argued that in *Richard III* "Shakespeare was writing altogether on the lines of Marlowe," while Charles Norman commented that "much that Shakespeare was to do is found in [Marlowe's] *Edward II* in epitome, and all of it is shadowed forth in verse not even he surpassed" (quoted in Pinksen 2008: 7). Although such scholars agree that Shakespeare's works surpassed Marlowe's, then if Shakespeare was Marlowe, this would be no more than a result of his development as a writer, something that is already evident from those works published under Marlowe's own name.

Stylometric analysis has provided some interesting, albeit conflicting, information on the authorship question. An independent study by T. C. Mendenhall in 1901 into the word length frequencies of different writers demonstrated that frequencies found in texts by Shakespeare and Marlowe matched each other exactly, while a recent analysis by T. Merriam of preferred common words shows eight Shakespeare plays to have an affinity with Marlowe's works that goes beyond mere influence. Not surprisingly, other studies, employing different stylometric tests, disagree with such findings. Future work in this field will undoubtedly contribute to the debate.

The Principal Alternative Authorship Candidates

Edward de Vere, Earl of Oxford

Edward de Vere, 17th Earl of Oxford, is currently the most popular of the alternative candidates. New books about de Vere are published regularly, and in both the United States and the United Kingdom, there are societies actively promoting his authorship.

Edward de Vere was born in about 1550 and was educated first privately, then at Cambridge and at Gray's Inn. Oxford traveled on the continent from 1575 to

1576, mainly in France and Italy. Oxford's first wife was Lord Burghley's daughter, Anne Cecil, with whom he had two daughters. Well-known scandals included his quarrel in 1579 with Sir Philip Sidney, which nearly resulted in a duel, and his affair in 1581 with Anne Vavasour, one of the queen's maids of honor.

In 1920 Thomas Looney identified Oxford as possessing the very background, learning, and experiences that he assumed the author of Shakespeare's works would have possessed, including classical learning and knowledge of law, music, Italy, and aristocratic habits and sports (Ogburn 1988). Some Oxfordians even claim Oxford as a leading figure in the Renaissance in England. Notably, two of his uncles, Lord Sheffield and Lord Surrey, were poets, while his uncle and former tutor Arthur Golding had translated Ovid's *Metamorphoses* (1567), on which Shakespeare's works frequently draw. Oxford was also patron of a company of players and was acknowledged in his lifetime as a poet and playwright. Oxford's extant poems under his own name, as well as his letters, are used to support claims for his authorship of Shakespeare's works (also, a quarter of the verses highlighted in a Bible said to have belonged to Oxford relate to biblical verses found in Shakespeare's plays). Moreover, Shakespeare's plays are said to be full of allusions to events and characters in Oxford's personal life, and Oxford's coat of arms as Viscount Bolebec shows a lion brandishing a broken spear. From 1586 onward, Oxford was awarded a grant of £1,000 per annum from Queen Elizabeth, the second highest such annuity, which Oxfordians claim was to provide him with the financial relief necessary to allow him to dedicate himself entirely to his writing, though the evidence is to the contrary.

Oxford's death in 1604 is said to coincide with the end of any new publications by Shakespeare, at least until the First Folio of 1623, which was dedicated, among others, to Oxford's son-in-law Philip Herbert, Earl of Montgomery. Although the conventional dating of Shakespeare's plays, based on circumstantial evidence, assigns several to the years after 1604, Oxfordians claim that the orthodox dating is flawed. In its place, they have suggested different dates for most of Shakespeare's plays, some, such as *A Midsummer Nights Dream*, as early as 1573 (well before Marlowe's revolution of blank verse drama).

Although Oxford did possess many of the qualities that anti-Stratfordians look for in the true Shakespeare, critics of the case for Oxford have argued that Oxfordians take many aspects of Oxford's life out of context in order to make them fit a hypothetical biography of Oxford as Shakespeare. Moreover, his public recognition and acclaim as a poet and playwright make it difficult to see why he had to resort to a pseudonym.

Christopher Marlowe

Willburn Gleason Zeigler first proposed Christopher Marlowe, the poet, playwright, and government spy, as an authorship candidate in 1895. Despite the comparative

soundness of the case made for him since then, Marlovians have been losing out to the Oxfordians in making their case better known.

Marlowe, the "professional" candidate for the authorship of Shakespeare's work, was born in 1564, the same year as Shakespeare, and came from a similar background, being the son of a Canterbury shoemaker. In 1578 Marlowe won a scholarship to the King's School, Canterbury, and in 1581 he matriculated at Corpus Christi College, Cambridge, also on a scholarship. While at Cambridge, Marlowe was recruited into Elizabeth I's secret service, which was then run by Sir Francis Walsingham and Lord Burghley, and for whom he worked under cover in Holland and possibly France. Marlowe's first ventures into poetry, most likely while still at Cambridge, were translations into English of Ovid's *Amores* and Lucan's *Pharsalia*. His first play, *Dido, Queen of Carthage*, was written in about 1586, and this was followed, probably in order, by *Tamburlaine the Great* (Parts 1 and 2), *Doctor Faustus, The Jew of Malta, Edward II*, and *The Massacre at Paris*, all by 1592, although some were only published after his apparent death on May 30, 1593. Marlowe also wrote some of the most popular poems of the period, *The Passionate Shepherd to His Love* and *Hero and Leander*. Some orthodox scholars also think he had a hand in Shakespeare's *Henry VI* plays and in *Titus Andronicus*. Marlowe's literary patron was Thomas Walsingham, and he was probably a member of Sir Walter Raleigh's circle.

It appears that Marlowe was far from conventional in his outlook on life, and on May 20, 1593, Marlowe was arrested at his patron's house following accusations of atheism made against him. He had also been implicated by his fellow playwright Thomas Kyd, possibly under torture. Marlowe was brought before the Privy Council but released on bail pending further inquiries. He is officially said to have died in a fight 10 days after his arrest and a few days before the first ever appearance in print of Shakespeare's name.

While Marlowe's apparent murder in 1593 should provide an even greater obstacle to his authorship claims than Oxford's early death in 1604, this is not necessarily so, as evidence of his death is far from conclusive (Farey 2005).

The official inquest into Marlowe's death reveals that Marlowe had dined and spent the day at a private house, along with three other men, who are now known to have been connected either to the Elizabethan secret service or to Marlowe's patron Thomas Walsingham. Toward the end of that day, Marlowe supposedly became violent in a dispute about who was to settle the bill (or "reckoning") and was killed in self-defense by one of his companions, Ingram Frazer. The inquest into his death, before the queen's coroner, took place two days later.

Could it really be a coincidence that when his fellow playwright Kyd had already been arrested and tortured, and when Marlowe is about to be reinterrogated, possibly under torture, that just at that time Marlowe is "killed" by an employee of his patron and his body buried quickly without any independent identification of his corpse? Even orthodox historians doubt the version of

events as recorded by the inquest, though they suggest that Marlowe was murdered in order to protect some of his associates (possibly Walsingham and Raleigh) from being implicated by him and to spare them embarrassment or even prosecution. Given the individuals involved in such a plot, however, such a scenario does not seem likely on close analysis. Instead, Marlovians suggest that Marlowe's death was faked, possibly with the aid of his patron, in order to allow him to be spirited away to safety, probably abroad. By pretending that Marlowe had been killed in self-defense, an inquest could be arranged, which would result in an official record of his death, should anyone ever claim in the future that they had seen Marlowe alive. If Marlowe were simply to be assassinated, no such complicated charade would have been required.

Marlovians consequently approach the authorship question very differently from other anti-Stratfordians. Quite independent of the Shakespeare authorship question, an analysis of Marlowe's death leads to the suggestion that he may still have been alive after 1593. This naturally leads to the question what a man of Marlowe's poetic genius would have done in these circumstances. Marlovians propose that Marlowe would have continued to write, but that his work would henceforth have to be published anonymously or under an assumed name, so as not to endanger those who had helped him escape from England. As it happens, works of Marlowe's quality and in his style (though one that matured and developed over the years) did indeed continue to appear in print and on stage, but under the name of William Shakespeare, arguably the William Shakespeare of Stratford, who allowed his name to be used and who acted as a play broker of the works supplied to him and his company by Marlowe via intermediaries.

Orthodox Shakespeare scholars have always acknowledged that without Marlowe's groundbreaking work, there would have been no Shakespeare. As stated by the Stratfordian Jonathan Bate (quoted in Pinksen 2008: 12), "Marlowe did come back from the dead after the Deptford stabbing; his ghost astonishes us as we read and hear the verse of Shakespeare." While the plays that are in Marlowe's name may show a less accomplished dramatic structure than those ascribed to Shakespeare, this is easily accounted for if one views any such difference as an indication of the development of Marlowe's work as a whole. And while many Stratfordian scholars like to claim that Marlowe's work differs considerably from Shakespeare's in its themes and preoccupations, most orthodox scholars of Marlowe make no such distinctions. For example, Marlowe's *Tamburlaine* is said to address, among other issues, "the division of women's loyalties in a patriarchal structure, a theme that Shakespeare would later explore at more length in *King Lear*," while *Tamburlaine*'s "combination of stoic fortitude with masochism and vulnerable openness . . . typifies the martial heroes of Shakespeare's Roman plays" (Chedgzoy in Cheney 2004: 249, 258).

In fact, orthodox scholars have long pointed out that Shakespeare's work echoes Marlowe's work throughout. John Bakeless (quoted in Pinksen 2008: 10)

argues that "Shakespeare quoted Marlowe or alludes to his plays repeatedly . . . practically the whole of Marlowe's work as it is know today." In *Hamlet*, Act 2, Scene 2, for example, Shakespeare alludes at length to Marlowe's *Dido, Queen of Carthage*. Given that Marlowe's plays contain numerous allusions to his own writings, is it not remarkable that when Shakespeare's plays allude to contemporary works they do so predominantly to Marlowe's? In *As You Like It* (Act 3, Scene 5) Shakespeare even quotes a line from Marlowe's poem *Hero and Leander*.

Shakespeare's work is generally said to culminate in *The Tempest*, whose main character, Prospero, a reworking of Marlowe's Faust character, spends 12 years stranded on an island. *The Tempest* was first published in the First Folio, where it appears before all other plays, as if it were an introduction to Shakespeare's work as a whole. No less an orthodox scholar than A. L. Rowse has commented: "Marlowe's historic achievement was to marry great poetry to the drama; his was the originating genius. William Shakespeare never forgot him; in his penultimate, valedictory play, *The Tempest*, he is still echoing Marlowe's phrases" (quoted in Pinksen 2008: 7).

One downside of the Marlovian theory, however, is that it does little to fill the biographical void, which many long for with a view to enhancing our understanding of the plays, as nothing is known of Marlowe after his supposed death. Whereas the proponents of Oxford and others write Shakespeare biographies based on what is already known about the lives of their candidate, Marlovians are only able to speculate about Marlowe's "posthumous" life from the generally accepted developmental stages of the plays and from the story found in the *Sonnets*.

Others

Out of the remaining more than 25 authorship candidates for which a case has been made at one time or another, arguably the most prominent are Francis Bacon, William Stanley (the Earl of Derby), Roger Manners (the Earl of Rutland), Sir Henry Neville, and one of several women, Mary Sidney Herbert. Mention should also be made of the Groupist theory, which suggests that Shakespeare's works were a collaborative effort by any number of individuals, possibly led by either Bacon or Oxford or Mary Sidney. That theory has some small merit, insofar as playwriting in the Elizabethan period was indeed at times a collaborative effort (even Stratfordians acknowledge that Shakespeare had worked with others, such as Fletcher). It is inconceivable, however, that a large number of individuals could have been responsible for Shakespeare's works and been able to keep that collaboration a secret.

Some Poems and Plays in the Light of the Authorship Debate

One reason that anti-Stratfordians give in defense of the authorship debate is that all of Shakespeare's works pose historical and literary problems that cannot easily

be explained if one assumes they were written by Shakespeare of Stratford. Some problems relating to two works, *Hamlet* and the *Sonnets*, are outlined below.

Hamlet

Hamlet, which was first performed by 1602, is considered to be Shakespeare's most autobiographical play. Among the many autobiographical aspects identified by Oxfordians is that just as Hamlet's mother Gertrude remarried with "unseemely haste," so did Oxford's mother. The character of Polonius is likewise said to be a caricature of Lord Burghley's, whose ward Oxford had been, while the verbose letters of Polonius make fun of Burghley's equally longwinded letters to the queen. Hamlet's killing of Polonius, who was spying on him from behind a tapestry, is apparently based on a similar event in which Oxford killed a man. Elizabeth I's resistance to Oxford's request for permission to travel abroad may be echoed in *Hamlet* (Act 1, Scene 2), when the king tells Hamlet:

> It is most retrograde to our desire;
> And we beseech you, bend you to remain
> Here, in the cheer and comfort of our eye,
> Our chiefest courtier, cousin, and our son. (Craig 1891)

Oxford was indeed Elizabeth's Lord Great Chamberlain and England's premier earl.

Marlovians, on the other hand, point to Hamlet's speech to the players at court (Act 2, Scene 2), which alludes to Marlowe's earliest play *Dido, Queen of Carthage*, which was not printed until 1594.

> I heard thee speak me a speech once, but it was never acted;
> Or, if it was, not above once;
> for the play, I remember, pleased not the million;
> 'twas caviare to the general: but it was . . .
> an excellent play, well digested in the scenes, set down with as much
> modesty as cunning. . . .
> One speech in it I chiefly loved;
> 'twas Æneas' tale to Dido. (Craig 1891)

Although Hamlet subsequently does not actually quote Marlowe's play, he echoes its content and style. What is intriguing about this passage is that Shakespeare (speaking through Hamlet) is not only familiar with the fact that *Dido* was not a success, but that he defends the play in terms only its author, Marlowe, would.

The Sonnets

Shake-speares Sonnets, a collection of 155 short poems first published in 1609, have not only baffled orthodox scholars, they are also arguably the key text in

the authorship debate. As orthodox scholars themselves used to claim, the *Sonnets* appear to be Shakespeare's most personal work. F. J. Furnivall commented in 1877, "no one can understand Shakespeare who does not hold that his Sonnets are autobiographical, and that they explain the depths of the soul of the Shakespeare who wrote the plays" (quoted in Pinksen 2008: 93). Similarly, F. S. Boas argued in 1896, "it is inconceivable that such intensity of passion as the Sonnets reveal should spring from no solid basis of fact" (quoted in Pinksen 2008: 93). Given the lack of any demonstrable link between Shakespeare of Stratford's life and the *Sonnets*, orthodox scholars nowadays consider the mystery of the *Sonnets* either unsolvable (as they dare not tackle the authorship issue) or consider them mere literary fancies. Unlike orthodox scholars, who have hence admitted defeat in their attempts to link the *Sonnets* with any known events or characters in the life of Shakespeare of Stratford, the proponents of alternative candidates have fared considerably better. It may be for this reason that Stratfordians writing on the authorship issue, such as Irvin Matus, avoid the issue of the *Sonnets* altogether.

It has been argued in support of Oxford's authorship of the *Sonnets* that in their dedication they are stated to be the work of "our ever-living poet," ever-living being a phrase used only of the dead, and that Oxford had died in 1604, five years before the *Sonnets* were published. Moreover, the "ever" in "ever-living" is considered a barely disguised rendering of E. Vere, Edward de Vere. A similar word play on Oxford's name can be found in Sonnet 76, which Oxfordians claim is the most direct cryptic clue to Oxford's authorship:

> Why write I still all one, ever the same,
> And keep invention in a noted weed,
> That every word doth almost tell my name,
> Showing their birth, and where they did proceed? (Craig 1891)

"Ever" the same, and "every word doth almost tell my name" are hence said to confirm Edward de Vere's authorship by playing on his name. Those lines are generally seen by anti-Stratfordians as confirmation that the author's identity was hidden behind a pseudonym (why would every word otherwise "almost" tell his name, when the author's name is supposedly featured on the title page?).

As for the first 17 *Sonnets*, there is almost universal agreement among orthodox scholars that these were composed as a commission from Lord Burghley with a view to encouraging Henry Wriothesley, the young Earl of Southampton, to marry. Both Oxfordians and Marlovians hence rightly wonder why Burghley would have approached the then apparently unknown and untried Shakespeare, when Oxford and Marlowe were well known to him and when both, according to their respective supporters, would have been the obvious choice to supply such poems?

Unlike the Oxfordians, proponents of Marlowe, however, are the only ones who have succeeded in finding a reasonably coherent and autobiographical explanation for the *Sonnets*. Marlovians have, for example, highlighted a sequence they call the sonnets of exile and anonymity, Sonnets 25–34, 36, 37, 39, 43–52, 56, 60, 61, 71, 72, 75, 76, 81, 97–99, 109, 113, 114, and 125. It is only in the light of the accusations leveled at Marlowe, and in the light of his faked death and exile, that those sonnets make any sense.

Only Marlowe, who was supposedly buried at Deptford, could write in Sonnet 72:

> My name be buried where my body is,
> And live no more to shame nor me nor you . . .,

while commenting in Sonnet 121:

> 'Tis better to be vile than vile esteem'd. . . .

These sentiments also appear in Sonnet 29:

> When in disgrace with fortune and men's eyes
> I all alone beweep my outcast state,
> And trouble deaf heaven with my bootless cries,
> And look upon myself, and curse my fate. . . . (Craig 1891)

There is nothing in Shakespeare of Stratford's documented life that even gives a hint of the disgrace that befell the writer of those sonnets.

Interestingly, Shakespeare's works are filled with characters who need to disguise themselves and who suffer exile and banishment. The author's personal anguish is probably reflected in Romeo's declaration (*Romeo and Juliet*, Act 3, Scene 3):

> banished is banish'd from the world,
> And world's exile is death. Then 'banished'
> Is death mis-term'd. (Craig 1891)

Lastly, Sonnet 81 offers important reflections on writing and fame, which are in stark contrast to what we know of Shakespeare of Stratford:

> From hence your memory death cannot take,
> Although in me each part will be forgotten.
> Your name from hence immortal life shall have,
> Though I, once gone, to all the world must die:
> The earth can yield me but a common grave,
> When you entombed in men's eyes shall lie.

> Your monument shall be my gentle verse,
> Which eyes not yet created shall o'er-read; (Craig 1891)

As Shakespeare of Stratford is known to have been concerned with his material and social status, he would presumably have delighted in the fame that his written work could bring him. The above lines, however, were clearly written by a man who, despite desiring such fame, knew that these poems would never truly be associated with him, even if the title page proclaimed them as *Shake-speare's Sonnets*.

The Cover-Up

Arguably the most important issues any alternative authorship theory must address are not only why it would have been necessary for the author of Shakespeare's works to hide his true identity, but also why his true identity had to *remain* hidden after the author's death. Considering that many anti-Stratfordians argue that it was known in the relevant circles that Shakespeare of Stratford was not the poet and playwright, it is even more surprising that there is no record, even in private papers, that reveals the possibly "open secret" of that authorship.

Anti-Stratfordians have explained the need for secrecy in several ways. Baconians argue that for personal and political reasons, Bacon was unable to acknowledge publicly his poems and plays. Oxfordians and the proponents of other noble candidates likewise claim that as a senior courtier Oxford could not have been seen writing for the public stage, though this does not explain why Oxford would have needed to publish poetical works such as *Venus and Adonis, Lucrece*, and the *Sonnets* under a pseudonym. Oxfordians also claim that any records of Oxford's authorship were destroyed on the orders of Lord Burghley, who had fallen out with his son-in-law over the latter's treatment of Burghley's daughter Anne.

All these arguments are difficult to sustain, however, and proponents of Bacon, Oxford, and others by far overstate the case for the need to keep their candidate's authorship a secret, not least posthumously. There was nothing sufficiently dangerous about Bacon's or any other aristocrat like Oxford's authorship (or even that of any woman) to merit such lasting secrecy.

Only in the case of Marlowe is it easy to see why he and his supporters had sufficient reason to keep his authorship a secret both during and after his lifetime. Considering that Marlowe was a fugitive from the law following his faked death in 1593, there were no circumstances in which Marlowe could ever have resurfaced under his own name, either in the British Isles or abroad. Even after his death, his supporters had to keep his erstwhile survival a secret, so as not to be implicated in his escape.

The plays and poems written by Marlowe after 1593 were consequently placed with publishers or with Shakespeare of Stratford, who effectively acted

as a play broker. This role would have given Shakespeare of Stratford the standing and means to acquire a stake in the Lord Chamberlain's Men, despite being a newcomer to the world of the London theater, while also allowing him time to pursue his various business interests in his hometown of Stratford. Even assuming that the Lord Chamberlain's Men suspected that the plays were not by Shakespeare of Stratford, they had nothing to gain from raising the issue.

Shakespeare of Stratford even continued to serve as a convenient front man after his death in 1616, and his posthumous role in the coverup was cemented with the 1623 publication of Shakespeare's collected plays in the First Folio.

Conclusion

The view that Shakespeare of Stratford was not the author of Shakespeare's works has gained considerably in popularity in recent decades, yet still struggles against the might of the orthodox Shakespeare industry. Inroads are being made, however. In 2007, for example, Brunel University in London became the first university to offer a postgraduate MA program in Shakespeare authorship studies. A "Declaration of Reasonable Doubt About the Identity of William Shakespeare," which was launched by the actors Derek Jacobi and Mark Rylance in 2007, has meanwhile attracted a number of academic signatories.

Nevertheless, anti-Stratfordian texts are still not being published by any of the major academic publishers or by any of the relevant academic journals. Peer review, meaning peer review by Stratfordians, presumably ensures that anti-Stratfordian articles, whatever their merits, are kept out of academic journals. Such a Stratfordian "conspiracy," however, denies anti-Stratfordian scholars the benefits that should come with peer review, which is an essential part of scholarly dialogue. As long as the authorship debate is denied a place within the academic mainstream, scholarship is stifled.

Admittedly, anti-Stratfordians have not made it easy for themselves on account of the untrained approach taken by many and the eccentricity of some of their arguments. Under these circumstances, it is easy for Stratfordians to mock anti-Stratfordian arguments. Irvin Matus, for example, in his *Shakespeare in Fact*, found many key claims against Shakespeare or in favor of one or another candidate to be seriously flawed, as they were based on a misreading of the evidence and on superficial understanding of Elizabethan politics, theater, literature, and so forth. Given that the majority of anti-Stratfordian writers are not experts in the Elizabethan period, and may not even be trained historians or literary scholars at that, this is perhaps not surprising, and should serve as a warning to anti-Stratfordians that their arguments have to be grounded in a thorough knowledge of the period. Stratfordians, however, are similarly guilty of ignoring the facts or subjecting them to unwarranted interpretations.

While anti-Stratfordians are in general agreement on the reasons why Shake-speare of Stratford could not have written Shakespeare's works, there is no agreement on an alternative authorship candidate. Given the prominence of the Oxfordian theory, Stratfordians have attempted to dismiss the authorship debate in general by focusing on the weaknesses of the arguments in favor of that par-ticular candidate, while largely ignoring the reasons that the debate exists at all.

Although the true identity of the author of William Shakespeare's works has not to date been established definitively, the authorship debate should not be excluded from academic debate. Orthodox scholars of Marlowe in particular should be open to exploring the case for Marlowe, which appears to provide the most straightforward answers to the key questions in the Shakespeare author-ship debate.

References and Further Reading

Anderson, Mark. *Shakespeare By Another Name: The Life of Edward de Vere, Earl of Oxford, the Man Who Was Shakespeare*. New York: Gotham Books, 2005.

Blumenfeld, Samuel L. *The Marlowe-Shakespeare Connection: A New Study of the Authorship Question*. London: McFarland, 2008.

Cheney, Patrick, ed. *Cambridge Companion to Christopher Marlowe*. Cambridge: Cambridge University Press, 2004.

Craig, W. J., ed. *The Complete Works of William Shakespeare*. Oxford: Clarendon, 1891.

Farey, Peter. "Marlowe's Sudden and Fearful End: Self-Defence, Murder, or Fake?" *Marlowe Society Research Journal* 2 (2005): 27–60.

Fields, Bertram. *Players: The Mysterious Identity of William Shakespeare*. New York: HarperCollins, 2005.

Greene, Robert. *Groatsworth of Wit*. London: 1592.

James, Brenda, and William D. Rubinstein. *The Truth Will Out: Unmasking the Real Shakespeare*. London: Pearson, 2005.

Matus, Irvin Leigh. *Shakespeare in Fact*. New York: Continuum, 1994.

Meres. *Palladis Tamia, or Wit's Treasury*. London, 1598.

Michell, John. *Who Wrote Shakespeare?* New York: Thames and Hudson, 1996.

Ogburn, Charlton. *The Mystery of William Shakespeare*. London: Penguin, 1988.

Pinksen, Daryl. *Marlowe's Ghost: The Blacklisting of the Man Who Was Shake-speare*. New York: iUniverse, 2008.

Shakespeare, William. *Mr. William Shakespeare's Comedies, Histories, and Tragedies, Published According to the True Originall Copies*. First Folio. London: Isaac Jaggard and Edward Blount, 1623.

Southworth, John. *Shakespeare the Player: A Life in the Theatre*. Stroud: Sutton Publishing, 2000.

Wraight, A. D. *The Story that the Sonnets Tell*. London: Adam Hart, 1994.

CON

The debate over whether or not William Shakespeare authored the great plays and other works that have traditionally been attributed to him is relatively recent in its origins. The first systematic challenges issued in the so-called authorship controversy date from the mid-19th century. Many researchers of various stripes and degrees of skill in presenting a case have since jumped into the debate, and dozens of candidates have been put forward as the "true" author of Shakespeare's works. Francis Bacon, the Earl of Oxford, Christopher Marlowe, the Earl of Rutland, the Earl of Derby, Walter Raleigh, Queen Elizabeth I, Francis Drake, Anne Whately, Cardinal Wolsey, many additional individuals and various groups of collaborators or conspirators have all been nominated by their proponents over the years.

The literature spawned by the controversy has grown to such unwieldy proportions that modern writers have adopted certain standard terms and spellings to avoid increasing the confusion. The author of the plays and sonnets, is, for example, referred to as "William Shakespeare," because that is the name that appears on the title pages of the published works. The man who was born in Stratford-upon-Avon, Warwickshire, England, in 1564, died there in 1616, and is now buried in Stratford's Holy Trinity Church under a slab near a monument to the achievements of Shakespeare is referred to as "William Shakespere" (the spelling that most often appears on legal documents that record the few surviving traces of his activities outside the theater). Elizabethan and Jacobean spelling was notoriously inconsistent.

Scholars, including the vast majority of professors of literature, who believe that Shakespeare wrote Shakespeare, are called "Stratfordians" and are sometimes referred to as "the orthodox." Those who believe that someone other than Shakespeare authored the works are called anti-Stratfordians, or heretics. Others who adhere to the theory that a group of authors collaborated to produce the works are called Groupists, some of whom argue for various combinations of noblemen, while others believe that Rosicrucians, Freemasons, or Jesuits are the responsible parties. Those who are convinced that Christopher Marlowe was the author are called Marlovians, supporters of Francis Bacon are Baconians, and proponents of the Earl of Oxford are Oxfordians. Bacon was the leading candidate throughout

much of the controversy's troubled and confused history, but his star has been considerably eclipsed since the 1920s by Oxford, the other leading nominee.

Anti-Stratfordians have spent many lifetimes in the compilation of data, consisting mostly of unexplained facts and inadequately supported speculations. Hundreds of books have been written in an attempt to discredit Shakespeare, including some by apparent cranks and others by earnest scholars who genuinely, but erroneously, believe in the validity of their positions. Understandably, given the welter of confusion among anti-Stratfordians, they have often been at odds with one another despite their shared interest in tearing down Shakespeare.

Principal Known Facts about Shakespeare's Life

The voluminous maze of research produced by anti-Stratfordians over the decades has resulted in some useful byproducts, if nothing else. New information about late 16th- and early 17th-century English literature has come to light. It is now apparent that we do not know much about Shakespeare's personal life, even though the biographical volume in the Yale Shakespeare Series reproduces 97 pages of legal, literary, or theatrical documents referring to him, according to Tucker Brooke (1926) in *Shakespeare of Stratford: A Handbook for Students*.

Briefly stated, the surviving records documenting Shakespeare's life indicate that he was born in Stratford (1564), married Anne Hathaway (1582), fathered a daughter named Susanna (1583) and twins named Judith and Hamnet (1585), and disappeared from the documentary record (during the so-called lost years) until he was 30, at which time he reappears on lists of actors who played for Queen Elizabeth I. Some jealous remarks, as recounted in Brooke's work, by rival playwright Robert Greene in 1592 (referring to him as an "upstart") suggest that the man from Stratford was established as a member of London's theatrical community by that date. By the age of 33 Shakespeare had acquired enough wealth to purchase prime real estate in Stratford, while simultaneously dodging taxes, moving from place to place, and getting involved in quarrels and lawsuits in London. In Stratford he made real estate deals and engaged in various business transactions. In London he bought shares in theatrical enterprises, acted in plays, and wrote poetry. Around 1598 the name Shakespeare or Shake-speare started appearing on the title pages of plays, some of which had been previously published without attribution. He went into semiretirement at Stratford between 1611 and 1613, died there at age 53, and was interred in the town's parish church (Holy Trinity Church) in 1616.

Contemporary allusions to Shakespeare usually refer to the literary works, not to the man himself. He was, however, known to have been a friend of fellow playwright Ben Jonson, who, according to Brooke, described him as follows: "He was, indeed, honest, and of an open and free nature." The most colorful

reference to him in the documentary record, again described by Brooke, appears in a diary written in 1601 by a law student named John Manningham, who relates a ribald theatrical anecdote. According to the story, the famous actor Richard Burbage was portraying King Richard III in Shakespeare's play when he attracted the admiration of a lady in the audience, who invited him to her bedroom later that evening. She instructed him to identify himself as "Richard III" when her servant answered the door. Shakespeare overheard this conversation, went to the lady's house, entered her darkened boudoir, and enjoyed her attentions before Burbage had a chance to do so. When Burbage arrived and had himself announced as "Richard III," Shakespeare sent back a message that William the Conqueror preceded Richard III. Manningham clarifies the humor by explaining that Shakespeare's name was William (Brooke 1926).

Convincing Proof of Shakespeare's Authorship

Internal evidence in Shakespeare's First Folio, comments by Shakespeare's friend Ben Jonson, and the inscription on the Shakespeare monument in Holy Trinity Church are among the keystones of the Stratfordian case, proving that Shakespeare was Shakespeare.

Jonson's poetic eulogy for Shakespeare in the 1623 First Folio edition of *Mr. William Shakespeare's Comedies, Histories, and Tragedies, Published According to the True Originall Copies*, clearly identifies the author of the plays as the man from Stratford (Shakespeare 1623). Baconians, Oxfordians, and other anti-Stratfordians have never convincingly addressed the eulogy's evidence or found a way to discredit it. Some have claimed that Jonson was surreptitiously praising some other writer, not Shakespeare, whose real identity he knew. The eulogy begins, "To the memory of my beloved, the author, Mr. William Shakespeare and what he hath left us." Jonson clearly identifies the author as Shakespeare of Stratford when, later in the poem, he refers to him as the "Sweet Swan of Avon." The Avon is a river, and Stratford is "Stratford-upon-Avon" (Shakespeare 1623). Anti-Stratfordians have ineffectually attempted to whittle away at this unambiguous evidence by pointing to supposed contradictions in the poem, often based on an incorrect interpretation of its grammar or on taking lines out of context. Much has been made, for example, of the line "Thou art a monument without a tomb." Anti-Stratfordians imply this means that Shakespeare was still alive in 1623 when the First Folio was published, even though Shakespeare had been buried at Stratford in 1616. The fuller context reads:

> Thou art a monument without a tomb,
> And art alive still, while thy book doth live,
> And we have wits to read and prayse to give. (Shakespeare 1623)

First Folio

Printed in a run of about 1,000 copies in 1623, the First Folio was the first collection of Shakespeare's plays. Although quartos (cheap pamphlets) of his works had been published before, *Mr. William Shakespeare's Comedies, Histories, and Tragedies* was the first near-complete anthology of his plays, lacking only *Pericles* and *The Two Noble Kinsmen*. Though it was published too late for Shakespeare himself to have had any input into it or the opportunity to correct the text, it was a professional enough endeavor (selling for about $200 in current U.S. dollars) that it's relied on as the early authoritative source for the bard's work. Many of the cruxes—errors deeper than typos, with no clear remedy—in the Shakespearean corpus date to this folio and continue to provide material for scholarly debate.

The context indicates that Jonson is referring to the frequently expressed sensibility that authors do not really die so long as their works are read and appreciated. The immortality of their thoughts figuratively extends their lives as well.

A similar sentiment is echoed by the commemorative poem Leonard Digges contributed to the prefatory matter in the First Folio, a poem that clearly links the author of the plays to Shakespeare, the man buried in Stratford. Digges wrote that when time had destroyed the author's monument in Stratford he would still be memorialized in print by the First Folio:

> Here we alive shall view thee soon. This Book,
> When brass and marble fade, shall make thee look
> Fresh to all ages. (Shakespeare 1623)

William Dugdale, an English antiquarian, made the first drawing of Shakespeare's monument in Holy Trinity Church in 1653. An engraving based on it appeared in Dugdale's 1656 book *Antiquities of Warwickshire*. The image shows a man resting his hands on what appears to be a cushion or sack, perhaps symbolic of a merchant's occupation. We don't know who paid for the monument (which was probably chiseled by stonemason Gerald Johnson) or who composed the lines that appear below the effigy, but it had to have been installed sometime between Shakespeare's death in 1616 and the allusion to the monument by Digges in 1623. The monument has been taken down several times over the years and repaired or repainted. The effigy now bears little resemblance to Dugdale's depiction, but his various books also contain images of other 17th-century antiquities that are quite different from what we see today. The monument as we know it shows Shakespeare with a goatee and an upturned moustache (as opposed to the drooping moustache in Dugdale's rendering) and different posture (Shakespeare 1623). The author's right hand is holding a pen, poised to write on paper. George Verne's drawing of the monument in 1737

does show the pen, so if Dugdale's depiction was ever accurate (which has never been proven), the effigy had been changed by the mid-18th century. Scott McCrea (2005) reproduces the images and explains the discrepancies in *The Case for Shakespeare: The End of the Authorship Question*.

The inscription on the monument in Holy Trinity Church compares Shakespeare to the greatest intellects and artists of past ages, a "Socrates in his genius, a Virgil in his art," and suggests that he has been reborn on "Olympus," home of the gods. The monument undoubtedly commemorates a writer (who, as we know, also functioned in his native town as a merchant). The inscription's closing lines read: "All that he hath writt / Leaves living art, but page, to serve his witt." The unambiguous implication is that Shakespeare of Stratford was an author of such accomplishment that all succeeding writers would only be fit to be his "pages" or servants.

There are plausible explanations for all questions that have been raised about the monument. There is no indication on the sculpture itself that it has ever been substantially altered. Even if the sculpture did once depict a businessman instead of a pen-holding author, a reasonable interpretation would be that the citizens of Stratford did not see many plays and did not hold playwrights in especially high esteem, preferring instead to honor Shakespeare as a local merchant and man of property. Despite clever but specious arguments, there is no convincing evidence that the monument contains encoded messages about Bacon or that it has been deviously altered at some point as part of a conspiracy to lend support to the Stratfordian position. The effigy is carved from a solid block of stone. According to John Michell (1996) in *Who Wrote Shakespeare?*, modern experts have found no signs of any substantial alterations.

The infamous, inelegant inscription on Shakespeare's burial slab on the floor of Holy Trinity Church, near the monument, has sometimes been cited as an example of the kind of crude verse that an uneducated man would write, in contrast to the elegant style associated with the author of the plays. The slab-stone reads:

> Good Frend for Jesvs Sake Forbeare,
> To Digg the Dvst Encloased Heare.
> Blese be ye Man yt Spares Thes Stones,
> And Cvrst be he yt Moves my Bones. (Roberts 1922)

If Shakespeare actually penned this snippet of doggerel verse, its crudity is easily explained. He knew his audience and, being a literary craftsman, was catering to it. He adapted his style to make the message more meaningful to an ignorant audience of parish clerks, sextons, and gravediggers, just the sort of people who would be in a position to someday empty his grave and reuse it. The intended audience would have better understood and been more receptive to plain speaking

and crude threats than to a more refined inscription. It was common practice for individually buried people to have their bones dug up and cast into a common charnel house, where they would be mixed with other human remains. If the dead were ever called to resurrection, it would be impossible under those circumstances for a man to be reborn with his original body intact.

Shakespeare's Education

Anti-Stratfordians often claim that there is an irreconcilable gap between the classical learning exhibited in Shakespeare's works and Shakespeare's supposed lack of education, emphasizing that several leading candidates, like Oxford and Bacon, had been well educated in the classics. But it is an unwarranted assumption that Shakespeare had no opportunity to acquire knowledge of classical literature. His father, an alderman in Stratford, was entitled to send his son to the local grammar school at no charge. The school's records haven't survived because they were destroyed by fire, but, as Marjorie Garber (2004) argues in her *Shakespeare After All*, grammar school education in Shakespeare's day provided a firm grounding in classical subjects. The Latin and Greek classics referenced in Shakespeare's works were an integral part of the grammar school curriculum. In fact, classical allusions were part of the "common store of knowledge" in the Elizabethan era

There is no reason to think that Shakespeare had no access to the books that Shakespeare knew. He may have had well-to-do associates who owned such books and were willing to share them. The fact that books are not detailed in Shakespeare's will does not mean that the man from Stratford didn't own them. Other highly literate people of the time died without bequeathing books. Shakespeare may have left whatever books he owned to his daughter, Susanna Hall, as part of the unspecified "goods" and "chattels" that she inherited. Or the will may originally have been supplemented by an attachment no longer extant, an inventory detailing property such as books.

Moreover, it is important to remember that some of the greatest English and American authors from various periods, for instance, Benjamin Franklin, Charles Dickens, and Mark Twain, were largely self-educated. The education argument put forward by anti-Stratfordians is deeply tainted by snobbery and elitism.

Shakespeare's Knowledge of the Law

Bacon was a lawyer, and Oxford had studied law at Gray's Inn. Anti-Stratfordians insist that Shakespeare's works contain so many references to the law that they had to have been written by someone with legal expertise. Nothing is known about Shakespeare's occupational training, legal or otherwise, but he was often exposed to lawsuits when his father was a Stratford town official. He may have assisted his father (who was illiterate) by writing up legal documents

or performing other law-related functions. Elizabethans and Jacobeans were highly litigious, and going to court was a common dispute-solving mechanism, even for small disagreements. Shakespeare himself, as the documents indicate, was involved in many lawsuits, and it would be surprising if he had not thereby absorbed some knowledge of legal terminology and procedures. Eric Sams (1995) wrote convincingly in his *The Real Shakespeare* that Shakespeare may have worked as a law clerk during the so-called lost years. Intriguingly, the renowned Folger Shakespeare Library in Washington, D.C., owns a copy of *Archaionomia*, a legal textbook published in 1568, which would have been a suitable handbook for an Elizabethan law clerk. The name "Wm. Shakespeare" is inscribed on the book's flyleaf.

Even if he had no personal legal expertise, Shakespeare could easily have called on the assistance of an acquaintance well versed in law to help work out the details of law-related dialogue. In any event, the knowledge of jurisprudence exhibited in the plays is not as sophisticated as anti-Stratfordians would like us to believe. Legal matters are not convincingly represented in every law-related scene in Shakespeare's works.

Shakespeare's Familiarity with Aristocratic Ways

Bacon and Oxford were both intimately familiar with higher circles of the nobility and court. But Shakespeare did not need to be a nobleman to write about kings, queens, and other highborn individuals. Commoners, servants, actors, and others had frequent chances to observe the nobility. All playwrights in Shakespeare's day incorporated noble characters in their productions because the play-going public demanded them and troupes liked to cater to the public's taste. Shakespeare, while acting with the highly regarded companies to which he belonged, played at aristocratic estates and appeared in productions at the royal court, affording ample opportunity to see and learn. His knowledge of aristocratic manners is more apparent in later plays than earlier ones, suggesting that his exposure to high society increased as his fame grew. But the knowledge of courtly protocol exhibited in Shakespeare's works is not by any means as sophisticated as anti-Stratfordians like to suggest.

Shakespeare probably had comfortable relationships with his "betters." He seems to have thought it possible that he could be accepted into higher social circles. He applied at one point for a coat of arms under the pretense (which was a genealogical fiction) that his lineage dated back to antiquity. In 1596 the Garter King of Arms at the Herald's Office drafted a coat of arms for Shakespeare showing a falcon supporting a spear. The proposed motto was "Non sanz Droict," meaning "not without right." Ben Jonson lampooned his friend's attempts at social climbing in *Every Man Out of His Humour*, a play in which the commoner Sogliardo is very puffed up about getting a new coat of arms, for which he paid

The draft of a Grant of Arms and Crest to John Shakespeare of Stratford-upon-Avon, father of playwright William Shakespeare. The document was created by Sir William Dethick, Garter King of Arms, in 1596 and is housed at the College of Arms in London. (Harry Todd/Getty Images)

handsomely. Carlo, a jester, compares the rampant boar in the crest to a "hog's cheek and puddings" served on a pewter dish, and Sir Puntarvolo, a knight, guffaws that the upstart's family motto should be "not without mustard." If Shakespeare was not already capable, at least to some extent, of rubbing elbows with the gentry, he would not have been emboldened to apply for a coat of arms, the granting of which would have bolstered his position in society. He was, we infer, in a position already to speak with and observe the nobility. Michell reproduces the proposed heraldic design with commentary in his book.

Anti-Stratfordians argue that Shakespeare's knowledge of aristocratic sports like falconry and tennis indicates that the author of the plays had to be an aristocrat. But falconry was ubiquitous in Shakespeare's day. He could easily have observed it firsthand and could have acquired knowledge of technical terms related to falconry from books or acquaintances. As Michell notes, Ben Jonson wrote that use of "hawking language," not hard to acquire, was an affectation of those who wanted to imitate the gentry's way of speaking. Many of the references in the plays to falconry involve caring for the birds, not the sport itself, and commoners, not nobles, would have been responsible for falcon tending. Tennis courts were also common in Shakespeare's day, and many commoners had opportunities to watch nobles play the game.

It is also important to bear in mind that Shakespeare's hometown was not the uncultured and unfrequented backwater that anti-Stratfordians seem to

believe, or would prefer to think, that it was. An archbishop of Canterbury and a mayor of London, for example, could both trace their origins to Stratford-Upon-Avon.

Foreign Languages and Distant Settings

Oxford, for example, traveled extensively on the continent, lived in Italy, and adopted Italian clothes and mannerisms. Other anti-Stratfordian candidates were probably better acquainted with foreign countries than Shakespeare would have been. Anti-Stratfordians allege that Shakespeare could not have known enough about foreign settings like Italy or Denmark to describe them as they are represented in the plays. In cosmopolitan London, however, Shakespeare must have frequently crossed paths and rubbed elbows with Italians, Frenchmen, demobilized soldiers and sailors, and a wide range of classes and nationalities, affording fertile opportunities for listening to stories and making inquiries about foreign lands. His access to travelers and foreign-born Londoners could have equipped him with whatever he needed to know. We must also take note that the sophistication of Shakespeare's geographic knowledge has been exaggerated by anti-Stratfordians. Plays set in Italy, for example, make no mention of the canals of Venice, the city's most salient characteristic.

Some have pointed to French dialogue in Shakespeare's plays and allege that Shakespeare's command of French could not have been adequate for this purpose. But French tutors were readily available in London, and the man from Stratford would have had many opportunities to learn. Christopher Mountjoy, his Huguenot landlord, was a transplant from France, and Shakespeare could have enlisted his help, or obtained other advice, for composing French-language scenes. Similarly, in London, he could have found occasional assistance with other languages he needed for a verisimilitude of scenes with foreign settings. Did Shakespeare have sufficient knowledge of military and nautical subjects and terminology to allow him to write convincingly about camps and ships, or battles and voyages? Shakespeare's plays contain indications that the author was familiar with battlefields, the soldier's life, shipboard jargon, storms at sea, and related matters. The streets of London in the 1590s were chockablock with former soldiers and sailors whom Shakespeare could easily have pumped for information after priming them with a pint or two of ale. Such informants could have provided firsthand accounts of battles on land or life at sea, military maneuvers and discipline, nautical vocabulary, ship's rigging, and other useful details.

The Nonexistence of Manuscripts, Letters, and Diaries

Anti-Stratfordians object that no documents in Shakespeare/Shakspeare's handwriting have survived aside from a few signatures on legal documents, including

three on several pages of Shakespeare's will. But the absence of handwritten material is not at all surprising. Very few diaries or letters written by commoners during the period in question have been preserved. Paper, handmade and usually imported, was expensive. After a paper document was no longer considered useful, for example, when a manuscript had been set in type, the paper was likely to be reused for some other purpose, such as stiffening the covers of a book. In Shakespeare's time the author of Shakespeare's plays had not yet risen to the apotheosis he later attained in the late 18th century. No one thought to preserve handwritten plays if they duplicated texts already available in print. Anti-Stratfordians who think it suspicious that handwritten manuscripts by Shakespeare are no longer extant are insufficiently familiar with basic facts about everyday life in the English Renaissance.

Similarly, the apparent lack of fanfare surrounding Shakespeare's death and burial is not mysterious. By 1616 he had retired and was living away from London. His plays were no longer the latest fashion. He had earned some degree of adulation, but was not in the 17th century the one-of-kind giant of literature that we now consider him to be. At the time of his death in Stratford he was a solid local citizen, a businessman who had written some well-received plays while living in the capital.

The Nature of Genius

Shakespeare is widely regarded as the greatest literary genius in English and perhaps the greatest writer of any place or time. If he had not been elevated to such heights in the literary canon, there would be no "authorship controversy." His deification has attracted iconoclasts. Other playwrights, like Ben Jonson, had little formal education, but no one considers it necessary to challenge their authorship. According to Harvard professor Marjorie Garber in *Shakespeare After All*, there may be a psychological need by some to attribute works of surpassing genius and timelessness to more than one person, to make the author of Shakespeare's works seem more like a god than a man. But, she confirms, there is no significant reason to doubt Shakespeare's authorship.

Shakespeare was an undisputed genius, and the workings of genius cannot always be explained. Shakespeare, the litigious Stratford property owner, had opportunities to learn things, meet people, and engage in experiences that would have been very useful to a playwright. Raw material lodged in his fertile imagination, fermented, and was in due course molded and crafted into brilliant works of literature. Shakespeare was not a man of normal ability. Few people are graced with such potential. Virginia Woolf, herself a great writer, never ceased to be amazed by his brilliant flights of words and ideas, comparing his writings in her *Collected Essays* to the incompletely understood process of musical composition. "From the echo of one word is born another word, for which

reason, perhaps, the play seems . . . to tremble perpetually on the brink of music" (1967: 28).

Such writers cannot be judged by the standards that apply to people of normal ability and attainment. Such minds have the ability to absorb information, to expand horizons, and to recast memories, observations, and pieces of book learning into timeless art. Like many great philosophers and artists, Shakespeare was better at posing fundamental questions about life than at providing answers. Yale professor A. D. Nuttall, in *Shakespeare the Thinker*, observed: "There is an individual human being behind the plays, but the man himself is elusive, endlessly mobile. . . . [The] plays are the product of a single, remarkable mind" (2007: 377).

Collaboration and Rewriting

Nevertheless, few experts believe that Shakespeare actually penned every word contained in Shakespeare's plays. It is well known that he (like all of his contemporary playwrights) collaborated on plays or rewrote the works of earlier dramatists to update them and make them more current, appealing, and marketable. A story from the late 17th century, for example, relates that the tragedy *Titus Andronicus*, attributed to Shakespeare and published in the First Folio, was actually an older play that Shakespeare rewrote. According to Michell, Edward Ravenscroft, who was revising the play in 1678 for a new staging, spoke with old men who had been involved in the London theater and were familiar with early 17th-century stage lore. "I have been told," wrote Ravenscroft, "by some anciently conversant with the stage that it was not originally his but brought by a private author to be acted, and he only gave some master-touches to one or two of the principal parts or characters." *Titus Andronicus* is not Shakespeare at his best, but Ravenscoft's information calls attention to what was a common and widely accepted practice in the Elizabethan and Jacobean theater world. As Brian Vickers (2002) argues in his *Shakespeare, Co-Author*, plays were passed from one author to another and were sometimes revised spontaneously during a run. They typically had to be wrestled into shape in only four to six weeks. Collaboration was a standard practice.

It is well established that Shakespeare borrowed from historical writings such as Plutarch's *Lives* and Holinshed's *Chronicles*. It is also plausible that he adapted a number of older plays that were part of his troupe's repertoire and that the plays, as improved by him, became famous. Such refurbished plays could subsequently have been published under his own name without mention of the less brilliant contributions of previous hands. In Shakespeare's time originality was not considered important. Audiences were not concerned about who wrote a play so long as they were entertained, or moved, and felt as they strolled away from the theater that they had gotten their money's worth. Shakespeare was a

brilliant rewrite artist, occasionally acquiring manuscripts of old or inferior plays and recasting them for dramatic effect, injecting psychological insight, rewriting humdrum dialogue, and transforming them into magnificent poetry and drama.

Anti-Stratfordian Fallacies

A researcher can sometimes become so obsessively focused on a theory, so singlemindedly absorbed, that inconclusive results seem like truth. Gaps in the record and logical inconsistencies fade away because of the power of belief, or desire to believe. Anti-Stratfordians have too frequently focused on evidence in isolation without perceiving the overall picture and coming to reasonable, balanced conclusions. They tend to be wishful thinkers who twist scattered bits and pieces of evidence. They are not adept at reading historical traces and have little understanding of historical context. We may well respond, as Gertrude did in *Hamlet*, Act 2, Scene 2, to the windbag Polonius, "More matter, with less art."

Some Baconians, for example, have resorted to crypto-analysis, a highly contrived line of investigation that scours Shakespearean texts for secret codes that supposedly reveal Bacon as the true author and purportedly explain his reasons for wanting anonymity. The "secret" is typically embedded in an anagram, which, ingeniously unscrambled, yields a Latin phrase replete with strained abbreviations (like "F. Baconis") and various cryptic and obscure references. Crypto-analysis is closely related to the Victorian era's fascination with occult phenomena. Such crypto-analytic techniques can distort texts in ways that seemingly, but speciously, support any proposition. The process has been thoroughly discredited by William and Elizabeth Friedman (1957), in *The Shakespearean Ciphers Examined*, and by others. Baconians have, over the years, squabbled among themselves and split into subgroups, several of which have strayed to the fringes of reason and coherence. Some have argued that Bacon was the rightful heir to the English throne, as revealed by crypto-analytic evidence that he was born of a clandestine marriage between the Earl of Leicester and Queen Elizabeth. Others are discontented with the theory that Bacon wrote only Shakespeare and advocate that he also produced the works of Montaigne, Cervantes, Spencer, and Marlowe.

Professors of English literature who specialize in Shakespearean studies overwhelmingly believe that Shakespeare, the man from Stratford, was Shakespeare, the author. Readers with balanced judgment and adequate preparation in history are not likely to be swayed by the convoluted arguments and unconvincing assertions brought to bear by anti-Stratfordians against Shakespeare's authorship.

Among Bacon's bona fide writings is his famous discussion of inductive reasoning (*The New Organon*, first published in 1620), wherein he pointed to the danger of reaching conclusions on the basis of personal experiences, individual predilections, and limited examples. He disparaged prejudices and preconceived attitudes (calling them "idols") and urged their abandonment (Bacon 2000).

Those who think that only a college-educated courtier like Oxford or Bacon could have written the works of Shakespeare fall into precisely the trap against which Bacon warned. Several anti-Stratfordian candidates, like Bacon and Oxford, were multifaceted and brilliant. Some were quintessential Renaissance men. But Stratford's William Shakespeare wrote Shakespeare's works.

References and Further Reading

Bacon, Francis. *The New Organon*. Edited by Lisa Jardine and Michael Silver-thorne. Cambridge: Cambridge University Press, 2000.

Brooke, Tucker. *Shakespeare of Stratford, a Handbook for Students*. New Haven, CT: Yale University Press, The Yale Shakespeare, 1926.

Dugdale, William. *The Antiquities of Warwickshire Illustrated; from Records, Leiger-Books, Manuscripts, Charters, Evidences, Tombes, and Armes: Beautified with Maps, Prospects, and Portraitures*. London: Thomas Warren, 1656.

Friedman, William, and Elizabeth Friedman. *The Shakespearean Ciphers Examined: An Analysis of Cryptographic Systems Used as Evidence that Some Author Other Than William Shakespeare Wrote the Plays Commonly Attributed to Him*. Cambridge: Cambridge University Press, 1957.

Garber, Marjorie. *Shakespeare After All*. New York: Anchor, 2004.

McCrea, Scott. *The Case for Shakespeare: The End of the Authorship Question*. Westport, CT: Praeger, 2005.

Michell, John. *Who Wrote Shakespeare?* New York: Thames and Hudson, 1996.

Nuttall, A. D. *Shakespeare the Thinker*. New Haven, CT: Yale University Press, 2007.

Roberts, Kate Louise, comp. *Hoyt's New Cyclopedia of Practical Quotations*. New York: Funk and Wagnalls, 1922.

Sams, Eric. *The Real Shakespeare: Retrieving the Early Years, 1564–1594*. New Haven, CT: Yale University Press, 1995.

Shakespeare, William. *Mr. William Shakespeare's Comedies, Histories, and Tragedies, Published According to the True Originall Copies*. First Folio. London: Isaac Jaggard and Edward Blount, 1623.

Vickers, Brian. *Shakespeare, Co-Author: A Historical Study of Five Collaborative Plays*. New York: Oxford University Press, 2002.

Woolf, Virginia. *Collected Essays*, Vol. I. New York: Harcourt, Brace, and World, 1967.

10

Galileo willfully violated the injunctions of the Inquisition and was thus guilty at his 1633 trial.

PRO   Joseph P. Byrne
CON   Arthur K. Steinberg

PRO

In the spring of 1633, the Florentine astronomer and mathematician Galileo Galilei was tried for heresy and convicted of "vehement suspicion of heresy" by the Holy Office of the Universal Inquisition of the Catholic Church in Rome. His error was the holding and teaching of four ideas that were central to the astronomical ideas of the Polish priest and astronomer Nicholas Copernicus (the "Copernican revolution"), who had died 90 years earlier: Earth is not at the center of the universe; the sun is at the center of the universe; the sun does not move; Earth both rotates on its axis and revolves around the sun in a circular path, along with the planets. As noted by Maurice Finocchiaro (1989) in *The Galileo Affair: A Documentary History*, despite a centuries-long tradition that denied these four claims, Catholic theological authorities had not made any definitive statements regarding them until 1616, when the censors of the Congregation for the Index of Prohibited Books suspended publication, sales, and even reading of Copernicus's key text *On the Revolutions*, and prohibited and condemned an Italian book that explained and defended Copernicus's ideas as physically true. In so doing, it also condemned "all other works in which the same is taught." Galileo, who was in Rome at the time and was a well-known proponent of Copernicus's ideas, received a stern, personal warning from the head of the Holy Office himself against holding or defending Copernicanism. Over the next decade, Galileo steered clear of writing about condemned Copernican ideas altogether.

The election of a new and seemingly more liberal pope, however, emboldened Galileo to write a long dialogue in which his characters discussed both Copernicanism and the accepted tradition, generally known as the Ptolemaic model. Though Galileo sought all the proper permissions from church authorities to publish his *Dialogue on the Two Chief World Systems*, flaws in the process and circumstances beyond Galileo's control allowed the book to slip through without proper review, and in 1632 it was published in Florence as he had written it. It was brought to the attention of the Holy Office in Rome,

225

which had the work reviewed by panels of theological experts twice. In both cases the experts agreed that the text clearly "taught" Copernican theory as physical truth and thus violated the 1616 ban. The authorities called Galileo to Rome from Florence to answer for his work, and the old and infirm scientist made the burdensome trip in January 1633. Though Galileo was treated with many courtesies, in the end he was forced to admit his "errors," formally condemn them, and agree never to hold or teach them again. He was to live out the remainder of his life—just under a decade—under house arrest in Florence.

From the day the trial ended, Galileo's experience with Catholic Church authorities has stood as an icon of the struggle between objective and truth-seeking science and religious authority that seeks to control knowledge and its dissemination for its own dubious purposes. Many myths surround the event—for example, that Galileo was physically tortured and that he defiantly muttered "*eppure si muove*" ("and yet it—the earth—moves") after formally denying that very statement. These have helped Galileo achieve the status of martyr for the cause of scientific truth in the face of ignorant and intolerant religious dogmatism, especially that based on biblical literalism. In our day, American society wrestles with many issues that have both a political and religious dimension, including the teaching of creationism or intelligent design theory as a viable alternative to the big bang theory and evolution; when a human fetus is deemed a person; and whether public schools should teach the Bible as a cultural text. In Galileo's day, the Catholic Church and its leaders constituted a powerful authority in Italy that wielded both spiritual and political influence, and its traditions and beliefs held sway throughout the Catholic world. Today, secular science and the secular state have displaced the church as, respectively, an authority on the natural world and the viable political power in it. The very notion of "heresy" and the institution of the Inquisition seem medieval and utterly irrelevant; the idea that one could be imprisoned, let alone tortured or executed, for holding or teaching a scientific theory is unthinkable in the post-Enlightenment world.

For the historically minded, however, the past must be viewed in its own light, and not through modernist lenses. For many historians and scientists, Galileo's trial and sentencing were not only a travesty by modern standards; they were incorrect by those of the 17th century. In short, Galileo was framed. If true, this doubly damns the Catholic Church as both intolerant and unscrupulous. Defenders of the church's actions have long since accepted that Galileo was later proven correct in his beliefs, but they contextualize the church's decision and defend the process by which it was reached. This debate hinges not on modern ideas of how truth about the natural world is obtained (the church long ago lost that debate), but rather on how one interprets historical documentary evidence, the heart and mind of Galileo, and the Holy Office's use of the term heresy.

Evidence that Galileo Was Guilty of Heresy

Christian authorities have used the term heresy since the very early church for ideas or claims that contradict official church teaching. The word itself is derived from the Greek word for choice. Christian doctrines (teachings) are many and most often deal with matters of faith and morals, though how this is construed has changed over time. Dogmas, teachings that must be believed by the faithful, include core Christian beliefs, such as the Incarnation of God in Jesus, and Jesus's historical death and Resurrection. These generally have long histories and were at some point formally defined by a pope or church council. The real problem for the church emerges when a heretic shares his or her ideas with others, especially when these contradict a firmly established dogma. Catholicism teaches that holding a false belief endangers one's own soul, but teaching it endangers those of the many others who might accept it, and the Bible itself is quite clear about the need to resist false teachers and teachings.

The Protestant reformation of the 16th century was sparked by the teachings of Martin Luther. These teachings resonated with many at the time but were condemned as heretical. Western Christianity rapidly unraveled, especially in northern Europe, resulting in a multitude of new Christian denominations that denied the legitimacy of Catholic authority. The Catholic Church was slow to react to these developments, but from the 1540s to 1563, the Council of Trent brought together Catholic bishops and theologians who reaffirmed the validity and centrality of Catholic authority, especially of the pope, and left to church authorities alone the right to interpret the meaning of Holy Scripture (the Bible). At the same time the pope established the Holy Office (1542), a judicial church court was established that tried suspected heretics and punished condemned heretics. Shortly after, the Index of Forbidden Books was reestablished as a means of identifying and censoring printed materials that contained heretical or offensive ideas. Its first edition was published in 1559, and the function of censoring and condemning was placed in the hands of a formal "Congregation" led by important Catholic cardinals in 1572. Neither the Holy Office nor the Index was brand new, but in the face of very successful challenges to church authority, they were reorganized and made devastatingly effective (if only among Catholics).

The Catholic Church had long accepted—though not as dogma—the model of the structure of the universe developed by the ancient Greeks and fine tuned by the second-century CE by natural philosopher Ptolemy. Earth held pride of place at the center of the universe and neither changed position nor rotated. The moon, sun, and planets occupied variously explained spherical "shells" and rotated around Earth, while a "sphere of fixed stars" created an outer boundary of the cosmos. Beneath the moon was the physical world of change, decay, and human life as normally experienced. Beyond the moon was a heavenly perfection without change, other than the perfect spherical rotations. And beyond

the fixed stars was heaven itself, to which Jesus literally ascended, according to scripture. Of course, the planets (literally wanderers) did not appear to rotate in a circular fashion, but rather erratically, causing Ptolemy to add many more circular motions to each planet. This "saved the appearance" as observed from Earth, but made a mess of perfect circular motion. Accepting the dissonance between theoretical perfection and observed complications, the church and its scholars made no serious attempts to reconcile the two. Indeed, theologians had found many places in the Bible where the inspired authors seem to support very clearly the static centrality of Earth and the revolution of the sun around it. The model also supported our own common experience (do we sense any motion?) as well as the physics developed by the Greek natural philosopher Aristotle, whose ideas had entered the Christian mainstream in the 13th century. One of his key ideas asserted that the "sublunar" earthly physics were qualitatively different from those of the unchanging heavens beyond the moon, where perfect circular motion was natural.

Nicholas Copernicus's model simplified the picture at a stroke by placing the sun at the center of the universe, and, for the most part, "saved appearances." It also challenged the entire traditional structure: the cosmology, the physics, the common experience, and, most dangerously, the church's interpretation of scripture. Because it was embedded in a book—*On the Revolutions* (1543)—that only appealed to (and could be understood by) a handful of mathematicians and astronomers, Catholic authorities largely ignored it. With the development of the telescope, however, and Galileo's use of it to study the heavens in the early 1600s, the situation changed.

Galileo Galilei was a Florentine—a subject of the Grand Duke of Tuscany. Though interested in painting and medicine, he studied mathematics at Pisa and became a tutor in Siena and Florence. He gained university appointments as a mathematician at Pisa in 1589 and Padua, near Venice, in 1592, where he taught until 1610. In the fall of 1609 he turned his rather crude 15-power (15x) telescope (and later a 20x) to the sky and began a series of (literally) world-shattering discoveries. The moon's surface was cratered, mountainous, and far from perfect. Jupiter has planets (moons) and revolves around the sun: so could Earth. Saturn, too, seemed to have two moons (its rings seen fuzzily at an angle), and Venus has phases like Earth's moon, as Copernicus had predicted. In 1611 Galileo visited Rome for a second time, demonstrating his telescope and sharing his thrilling observations with all comers. By 1615 sunspots had become evident, further eroding celestial perfection. Galileo and his telescope were redefining the heavens.

But the heavens were the church's domain. Some critics attacked the telescope as a machine that distorted the vision and created false impressions. Others, such as the Florentine philosopher Ludovico delle Colombe, lined up the scriptural passages that contradicted Copernicus's conclusions, and the interpretation of whose meaning lay with the church and not natural philosophy.

Ecclesiastes 1:5 teaches "The sun rises and sets, and returns to its place"; I Chronicles 16:30 states "God made the orb immobile"; and Psalm 104:5 asserts "You [God] fixed the Earth on its foundations." When the Bible stated that God allowed the Hebrew leader Joshua to "stop the sun" in the sky to prolong a battle, the church was inclined to take it literally, or as interpreted by the early fathers of church theology, a stance reinforced by the Council of Trent.

Galileo responded to these critics with a scornful attitude and a clear vision of his own regarding scriptural authority and human sense experience. In a famous and widely circulated letter to his old student Benedetto Castelli, dated December 21, 1613, Galileo explained that biblical authors actually used simple language of common experience and imagery to teach and ordain those things that are necessary for human salvation. Scripture was always true, but not necessarily literally. When human sense experience, such as his own observations or Copernicus's conclusions, clearly contradicts the literal sense of scripture, the Bible must be interpreted less literally, since scriptural meaning and sense experience cannot contradict one another. Though this stance seems common-sensical today, even to most Christians, Galileo was treading on thin ice by poking his nose into theological matters. Indeed, the letter was submitted to the Holy Office for review, though no fault was found at the time.

The head of the Holy Office at the time was Cardinal Robert Bellarmine, a Jesuit scholar and champion of Catholic orthodoxy. On April 12, 1615, he wrote a letter to the Neapolitan Carmelite theologian Paolo Foscarini, whose very pro-Copernican book had come to his attention. In the letter, Bellarmine

Martin Luther

Martin Luther (1483–1546) was a German Catholic monk who became an early reformer and schismatic in the earliest days of the Protestant reformation. Though it is unlikely he set out to have as enormous an effect as he did, he insisted that salvation was attained solely through faith (essentially denying the necessity of the church as a temporal power). When, like many others, he objected to the church's practice of selling indulgences to pay to rebuild Saint Peter's Basilica, he wrote a lengthy dispute (his '95 theses'), which he presented to his archbishop. Because of 41 heretical statements identified in his theses and his refusal to recant them, he was excommunicated from the church—which succeeded only in forcing him to seek support outside it, among the German princes.

Luther's new sect set forth many principles that have remained important among Protestants, including a distrust of the institution of the church as a mediator between the laity and the divine, the ability of clergy to marry, and the use of the vernacular. In his time, he also made vitriolic attacks against the pope, the anabaptists, and the Jews, the latter of which are sometimes blamed for inspiring much of the anti-Semitism in Germany.

stated that he was willing to accept that the Copernican model indeed saved appearances and had predictive value as a *hypothetical* model. But he was utterly unwilling to accept the Copernican model as physically true unless it were proven indisputably true. He and Galileo agreed that all of the evidence to date was only that, evidence, and was by no means *proof* of any of the four Copernican claims. In the absence of such proof, the church had to retain its position. Should undisputable proof emerge, then the church would need to revisit its interpretations and reformulate them in light of agreed upon facts. Scripture could not err, but an interpretation—even one commonly held by the church—could. Bellarmine believed, however, that astronomy could never prove the Copernican model to be physically true. He was also convinced that the subject matter of Copernicanism—cosmology—was indeed a matter of Christian faith, discussed in scripture, and thus properly left to the church and its agents, not to natural philosophy (science) and its acolytes. Direct contradictions to the church's teachings, however seductive or even correct Copernicanism might have seemed, at least verged on being heretical.

Historical Proofs Supporting Galileo's Guilt

In a letter of January 12, 1615, to Galileo from his friend Prince Cesi, the prince noted that Bellarmine had told him personally that Copernicanism is "heretical" and without a doubt contradicts scripture properly interpreted. On February 8, 1615, another friend also passed along good advice: Galileo was warned again to stay away from theological or biblical issues, this time by the sympathetic and powerful Florentine cardinal Maffeo Barberini. Despite these admonitions, in June Galileo resurrected his letter to Castelli on scripture and science and readdressed it to the Florentine grand duchess Christine of Lorraine. He circulated it widely, but did not publish it, lest it fall to the censors of the Index. Galileo followed this with a visit to Rome late in the year to defend himself against "rivals" and to use his considerable powers of persuasion to further the cause of Copernicanism and convince church authorities of its validity. Without
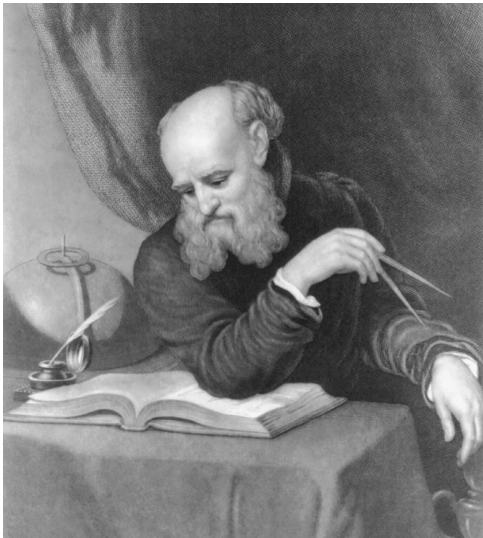
A leading figure of the scientific revolution, Galileo helped usher in the modern era by laying the foundations of mechanics and experimental physics. (Library of Congress)

doubt he firmly accepted Copernicanism and in good faith believed strongly that the church was better served by accepting than resisting it. Unbeknown to him, he had in fact been denounced to the Holy Office. Pressure on the pope and others to allow the free discussion of Copernicanism backfired, and the matter of its defense as fact was sent to the Holy Office.

While Galileo was making the rounds of prominent Romans, seeking their support, Bellarmine had Copernicus's *On the Revolutions*, as well as several other pro-Copernican works, including Foscarini's, scrutinized by theological advisers to the Congregation of the Index. Their decision, announced to the Congregation on February 24, was unambiguous and unanimous. As to the sun's centrality and lack of motion, the notion was absurd and foolish according to natural philosophy and "formally heretical," because it contradicted scripture, both literally and as interpreted by the church.

That Earth moved and was not at the center of the cosmos was equally foolish and absurd philosophically, and at least "erroneous" according to Catholic doctrine. On March 5, the Congregation announced its decision in a formal decree that was ratified by the Holy Office with the pope presiding: the books in question contained "various heresies and errors." This amounted to a formal condemnation of the ideas that only a pope (or perhaps a Church Council) could overrule, because popes ratified actions of the Holy Office. It was thus not an infallible (incapable of being in error) statement by which the church was eternally bound. Foscarini's work "and all other books that teach the same" (static sun at the center and rotating Earth revolving around it) were prohibited and condemned. Copernicus's own work was only subjected to "suspension until corrected," a process finished a few years later and imposed on all existing copies. It remained much longer, however, on the Index itself (until at least 1681), even in corrected form. The corrections that rehabilitated *On the Revolutions* imply that the censors were willing to accept Copernicus's ideas as hypotheses, or one theoretical way to understand the structure of the cosmos, but not to allow his model to be posited as the true, physical form of the cosmos, the mistake for which Foscarini's work (and later Galileo) suffered.

Though Galileo's name and writings were not formally linked to this decision (apparently thanks to the intervention of Cardinal Barberini), he was by no means immune to its effects. Two days after the advisers rendered their opinion Galileo was summoned to a meeting with Cardinal Bellarmine, whom he knew socially. What one understands about what happened at this meeting in many ways defines whether one accepts the validity of the judgment against Galileo in 1633. In dispute is what specifically was told to Galileo and by whom and the binding effect of the words. Subsequent to reception of the advisers' report, Pope Paul V, who was acquainted with Galileo and generally sympathetic to him, instructed Bellarmine to warn the astronomer of the impending action of the Congregation of the Index and to admonish him "to abandon these opinions."

Were Galileo openly to refuse, the Dominican father commissioner of the Holy Office, Michelangelo Segizzi (sometimes spelled Seghizzi), was to formalize the admonition into a legally binding injunction, delivered before witnesses and recorded by a notary, against his teaching, defending, or even "discussing," the condemned propositions. If this failed to elicit Galileo's agreement, he was to be imprisoned. These were the instructions, according to Finocchiaro in *The Galileo Affair*, as recorded in the minutes of the Inquisition on February 25, 1616.

The following day the cardinal met with Galileo and presented the pope's own warning; and then things get fuzzy. Did Galileo dutifully (and wisely) accept the cardinal's message, or did he refuse and trigger Segizzi and his formal injunction? According to the minutes of the next meeting of the Inquisition, Bellarmine reported that Galileo "acquiesced when warned." Indeed, in a certificate specially prepared at Galileo's request and dated May 26, 1616, Cardinal Bellarmine denied swirling accusations that Galileo had formally abjured his Copernican positions or had suffered any physical or spiritual punishments or penances at or as a result of the meeting of February 26. Bellarmine indicated that the astronomer had been warned against *defending or holding* the positions that had been determined to be contrary to the Bible. Neither the minutes nor the certificate mention either Segizzi or a formal injunction that would have been delivered had Galileo not cooperated with Bellarmine. A third document, however, does. It is the "Special Injunction" dated February 26 and proven (by paper and handwriting analysis) over a century ago to be from the time in question and not a later forgery. It purports to record Segizzi's actions immediately following Bellarmine's warning. It implies that Galileo did not have the time to accept or reject the cardinal's admonition before Segizzi laid out the formal injunction against holding, teaching, or defending the offensive ideas "orally or in writing," under threat of imprisonment and formal proceedings by the Holy Office.

The Special Injunction is dated and lists all who were present at this informal meeting, but the problem is that no one signed it. It was thus never a legal document. One explanation is that it was prepared beforehand in expectation of Galileo's negative response and, since Galileo acquiesced and it was not needed, it was never signed. But then, why was it not discarded? And why did its author portray Segizzi as failing to wait for Galileo's initial response and instead "thereafter, indeed immediately" confront the Florentine? Since this action would have been in direct contravention of the pope's instructions and may have had the effect of negating the injunction itself, and puts Segizzi in a bad light, it is very unlikely that it was written beforehand.

If a true memorial, then why did no one sign it? The most logical explanation is that Segizzi indeed blurted out his dire message, knowing that his formal injunction with its clear restrictions, presented on behalf of the pope and Holy Office, carried far more weight than the old cardinal's advice and warning. Segizzi was no amateur in these matters, and he must have calculated that since

the meeting between Bellarmine and Galileo was not a formal procedure, his intemperate move would not have affected the legal force of the injunction. Also, his action prompted the written Special Injunction in which his message was recorded. Even so, he had acted against the pope's expressed wishes, and it takes little imagination to see why those present would not have signed off on it. The whole point had been *not* to get Galileo in trouble if possible, to skirt the formalities embodied in the injunction itself. Bellarmine clearly wanted to bury Segizzi's action by leaving it out of both the following weeks' minutes and the certificate he provided to Galileo. This may have been a way of protecting Segizzi against papal wrath or out of Bellarmine's own at the Dominican's rashness.

Either way, Galileo left that meeting with the clear message that Copernicanism was off limits, a message echoed with the formal verdicts against Copernicus and Fracastorio "and others that taught the same" of March 5. One clear sign that Galileo took very seriously whatever he was told is the fact that he avoided writing about Copernicanism throughout the following eight years. During that time, Bellarmine and Pope Paul died (both in 1621); Segizzi was exiled to a post in Lodi, Italy; and the Florentine cardinal Maffeo Barberini, a close friend of Galileo's, was elected Pope Urban VIII (1623). Galileo had also made many new enemies by engaging the prominent Jesuit Orazio Grassi in a debate over a comet, even dedicating his brilliant manifesto, *The Assayer*, to the grateful new pope in 1624.

Heartened by his trip to Rome in the same year, Galileo began pushing the envelope on Copernicanism by writing his "Reply to [Francesco] Ingoli." Ingoli had written against Copernicanism in 1616 and carried out many of the revisions that had rehabilitated *On the Revolutions*. "Reply" is an attempt to demonstrate his firm understanding of Copernicus's ideas "which I then [before the condemnation] considered true," and his ability to parry Ingoli's arguments against them. "I was completely wrong in this belief of mine," he goes on, and notes that the ideas in question were "declared repugnant and suspect," and he has no intention of defending them as true, simply of rebutting Ingoli's objections as an intellectual exercise. At the end, he alludes to a work in progress on the tides in which he takes the motion of Earth as an hypothesis and deals with the whole debate. Galileo was convinced that the tides were caused by the motion of Earth and considered this his strongest evidence for that motion.

Though he was dissuaded from writing on tides, his next work took on the whole Copernican system and earned him the label heretic. *Dialogue Concerning the Two Chief World Systems* (Ptolemaic and Copernican) is a brilliant discussion of the respective merits of the two models. The fictional Aristotelian "professor" Simplicio (very close to "Simpleton" in Italian) defends the status quo, and does so quite clumsily. The nonfictional Florentine Filippo Salviati explains Copernicus's ideas and undermines Simplicio, and the actual Venetian nobleman Giovanfrancesco Sagredo plays host and generally concurs with the

reasonableness of Salviati's positions. It is nearly impossible to read, or even read about this work, without being convinced that it is a thinly disguised explanation and defense of that which was neither to be held nor defended, especially by Galileo. Indeed, it is peppered with spring-spined bows to the hypothetical nature of the theories Salviati presents. But the overall effect is clear. Galileo and his manuscript received a very positive reception in Rome in 1630, including an audience with Pope Urban around May 20. Raffaelo Visconti reviewed the work on behalf of the censor's office and concluded that it was too assertive of the theories and he needed to reduce them to hypotheticals. Galileo returned to Florence to make changes and prepare the text for printing.

The death of his patron in Rome, Prince Cesi, and an outbreak of plague reduced Galileo's ability to communicate with the Eternal City, its censors, and press. The chief censor, the Florentine friend of Galileo, Niccolò Riccardi, had supported Galileo's project and agreed to review the proof sheets as they came off the Roman press. But Galileo could not get the final manuscript to Rome and sought permission to have it reviewed and approved in Florence by the local ecclesiastical authorities. Under duress, Riccardi granted this permission in the summer of 1631, but he stipulated that the text had to avoid theology and make it clear that its purpose was to show that Roman authorities were aware of even the best arguments in favor of Copernicanism (and condemned it out of conviction rather than ignorance). Further, he had to make it clear that he took Copernicanism to be a hypothesis and not a true description of physical reality; and finally he had to include a statement reflecting the belief by the pope that God could act in any of numerous ways in shaping and maintaining the cosmos as long as they were not logically impossible. In the end, Galileo placed the pope's sentiment in the mouth of Simplicio, making it appear to be just another weak argument (though Salviati obligingly concurs). Simplicio's statement that one would have to be "excessively bold" to accept only one possible explanation undermines the church's position as much as it does Copernicus's.

When eight presentation copies finally appeared in Rome in the fall of 1632, Riccardi wrote to the Florentine authorities that the finished product contained many "unacceptable" things, that distribution would be blocked, and that the Congregation of the Index would review it for formal errors. A commission made up of Riccardi, the papal theologian Agostino Oreggi, and Jesuit Melchior Inchofer formally reviewed Galileo's *Dialogue* and reported to the Congregation and the pope in September. In addition to findings that he "may" have "asserted absolutely" the heliocentric theory and "may" have attributed falsely the tidal action to the movement of Earth, they found that he "may have been deceitfully silent" about the 1616 injunction that forbade him outright from dealing with Copernicanism. They had discovered the unsigned document in Galileo's file and brought it to the attention of the pope and the Congregation. Though modern commentators generally consider this to be without legal

standing, even in the judicial proceedings of the Holy Office, it is clear that it was henceforth key to the prosecution of Galileo.

The astronomer was called to Rome to appear before the Holy Office, which he did for the first time on April 12, 1633. Galileo produced his certificate from Bellarmine and denied recalling the more formal injunction that the court produced. He also denied vehemently that he supported the case for Copernicanism in the *Dialogue*, claiming—quite disingenuously—that the *Dialogue* actually supported the status quo. The court took the last claim seriously enough to have Oreggi, Inchofer, and theologian Zaccaria Pasqualigo review the book and report to the court. On April 12 each submitted his report. Oreggi briefly concluded that the *Dialogue*'s text was clear: Galileo held and defended Copernican ideas. At greater length Pasqualigo concluded that Galileo clearly taught and defended Copernican ideas and was suspect of holding them. Inchofer produced a longer and more detailed analysis, including 27 specific instances by which he demonstrates that Galileo teaches, defends, and holds the opinion of Earth's movement. Defending or holding violated Bellarmine's warning and Galileo's promise; defending, teaching, and holding violated the injunction.

Unlike America's presumption of innocence until guilt is proven, continental systems, including canon law, under which the Inquisition operated, presumed guilt and provided the accused an opportunity to defend himself. Determination of exactly what he was guilty of was the purpose of the "trial" in the spring of 1633. "Vehement suspicion of heresy" falls between the more serious findings of "formal heresy" and "strong suspicion of heresy" and the lesser findings of "mild suspicion of heresy" and various levels of less serious transgressions: erroneous, scandalous, temerarious, and dangerous. As the court made clear in its verdict, Galileo had chosen to hold, teach, and defend propositions he knew had been condemned by the Holy Office and had disobeyed both Bellarmine's warning and the injunction, the evidence for which the court accepted. These were matters less of doctrinal heresy than of what French historian Léon Garzend called "inquisitional heresy": to deny propositions that scripture clearly supports, or that church doctrine embodies but may not define, or that lower ecclesiastical authorities proclaim. In the end, Galileo's transgression was a matter of disobedience and flouting of authority as much as it was attachment to a condemned set of ideas.

Historical Interpretations Supporting Galileo's Guilt

In his *Retrying Galileo*, Maurice Finocchiaro traces the history of the Galileo controversy from his own day to the "rehabilitation" of the pioneer scientist by Pope John Paul II in 1992. Supporters of Galileo's guilt and the validity of the court's verdict have always been on the defensive and have almost always been Catholic apologists. By the 18th century, the major issues surrounding the trial had faded into obscurity, and defending Galileo's condemnation seemed pointless. The first

fully reasoned and documented defense came in a series of lectures in 1792 by the Jesuit Girolamo Tiraboschi. He depicted Galileo as rash and morally flawed, while deeming the church correct to defend the literal interpretation of scripture in the absence of clear proof to the contrary. Scientists, Protestants, secular historians, and philosophers and artists have eagerly attacked not only the verdict but also the basis for the process itself: church authority, papal power, the Inquisition, censorship, the very notion of "heresy." Historically, few had access to the church records necessary to draw a knowledgeable conclusion, and fewer still were (are) sympathetic to the outcome.

The simplest questions to ask are: Were Copernicus's ideas condemned in 1616 and did Galileo continue to hold them as true thereafter? While "the church" never did condemn heliocentrism formally in a papal or consiliar decree, the Holy Office did, and it would be that body that would try those who chose to hold and defend it, as Bellarmine's message made clear. Until almost the end of his ordeals in 1633, Galileo insisted that he had not "held or defended" Copernicanism since the 1616 condemnation, despite all evidence to the contrary. Galileo had gambled with his *Dialogue* and lost; though posterity is all the richer for it.

Conclusion

After 1633, Catholic authorities proceeded against no other Copernican writings or authors. In 1822 the Holy Office formally ruled that Roman Catholics could in good conscience believe in Earth's motion as physical fact. In 1835 both Copernicus's and Galileo's works were taken off the Index (though all of Galileo's works had been published as early as 1744). By the early 19th century, the physical proofs required by Bellarmine or Barberini had been provided by astronomers, and the Holy Office acted accordingly. Galileo had challenged the authority of the church not only to describe the heavens, but also to interpret scripture (in his Letter to Castelli/Grand Duchess Christina). He insisted that religious authority had to give way before the "facts" suggested (not proved) by the brand new and largely untried methodology of science. With Francis Bacon, Galileo was the founder of scientific method, and yet he firmly expected the millennium-and-a-half-old institution, founded and defined by divine revelation, to bow before his interpretation of his personal experiences, observations, and conclusions. He stood at the head of the centuries-long process that gave intellectual and cultural pride of place to the very methodology he stubbornly pressed. The centralized religious authority Galileo sought to inform ended up being trampled by the scientific revolution he helped spur and the Enlightenment that formally embraced and spread the revolution. Separating natural philosophy (science) from religion and theology was as radical in 1630 as separating church from state was in late 18th-century America. Significantly, echoes of both controversies reverberate in our ongoing national dialogue.

References and Further Reading

Biagioli, Mario. *Galileo Courtier: The Practice of Science in the Culture of Absolutism*. Chicago: University of Chicago Press, 1993.

Fantoli, Anibale. *Galileo for Copernicanism and for the Church*, 2nd ed. Notre Dame, IN: Notre Dame University Press, 1996.

Finocchiaro, Maurice. *The Galileo Affair: A Documentary History*. Berkeley: University of California Press, 1989.

Finocchiaro, Maurice. *Retrying Galileo: 1633–1992*. Berkeley: University of California Press, 2005.

Galilei, Galileo. *Dialogue Concerning the Two Chief World Systems*. Translated by Stillman Drake. Westport, CT: Greenwood, 2000.

Lindberg, David C., and Ronald Numbers, eds. *God and Nature: Historical Essays on the Encounter between Christianity and Science*. Berkeley: University of California Press, 1986.

McMullin, Ernan, ed. *The Church and Galileo*. Notre Dame, IN: Notre Dame University Press, 2005.

Olson, Richard G. *From Science and Religion, 1450–1900: Copernicus to Darwin*. Baltimore: Johns Hopkins University Press, 2004.

Poupard, Paul, ed. *Galileo Galilei: Toward a Resolution of 350 Years of Debate*. Trans. I. Campbell. Pittsburgh: Duquesne University Press, 1987.

Redondi, Pietro. *Galileo Heretic*. Princeton, NJ: Princeton University Press, 1989.

Roland, Wade. *Galilieo's Mistake: A New Look at the Epic Confrontation between Galileo and the Church*. New York: Arcade, 2001.

Sharrat, Michael. *Galileo, Decisive Innovator*. New York: Cambridge University Press, 1996.

Shea, William R., and Mariano Artigas. *Galileo in Rome*. New York: Oxford University Press, 2003.

Wallace, William, ed. *Reinterpreting Galileo*. Washington, DC: Catholic University of America Press, 1986.

Westfall, Richard S. *Essays on the Trial of Galileo*. Notre Dame, IN: Vatican Observatory, 1989.

CON

Catholic hostility toward "heretical" ideas was in part a reaction to forces the church could not control. The international threat from the rise of Protestantism, the northern European competitor, and Islam in the Mediterranean challenged a theological monopoly, which had existed for 1,500 years, the beginning of the Christian era.

Martin Luther, a former Catholic cleric, opposed the concept that one's faith alone would justify one's religious beliefs, a problem resolved by purchasing indulgences. At the same time, the church was confronted with the end of feudalism and the rise of nation–states, each of which became their own center of power. States seized the moment to declare their religious independence of Rome; they no longer sent money to the papacy and created their interpretations of religious beliefs that were inimical to Rome.

At the same time, another "ism" challenged Rome from the south. Muslim movements in the Mediterranean Sea, the Iberian Peninsula, North Africa, and portions of Central Europe saw the authority of the church recede. Muslim galleys captured Christian galleys, enslaving those who would not convert to Islam, interfered with trade, and threatened portions of the Italian peninsula. Muslim successes on Rhodes and the accomplishments at Malta and other islands caused some concern, but not enough for Christian Europe to oppose this expansion.

These territorial losses contributed to the hardening of Catholic dogma and the need to control the religious imperative: papal control and tightening of the sacred scriptures. Since the papacy was the beneficiary of "truth," theological interpretation became the weapon of choice within which to draw or limit the confines of knowledge. The church viewed rudimentary science as part of philosophy and as subject to the same restrictions.

Renaissance ideals and free thought expanded the borders of knowledge to the detriment of the church. Freethinking challenged the ancients' concepts of the universe and all that they had decreed as the truth. The rising challenges to biology, chemistry, and physics were only the entryway along a path hoping to understand why. Galileo was one of the first soldiers in the struggle to learn, who removed intellectual shackles from the intellects and offered his discoveries to society. Their successes began to challenge the administration of canon and civil law in papal-controlled regions. With the church functioning within the parameters of divine revelation, the fearful pope had to deal with heretics who defied accepted church doctrine.

Galileo's intellectual development occurred at a time when innovation and invention competed with papal fear, a vehicle used to circumvent the intellect and progress. He was born in Pisa, Italy, on February 15, 1564, the unwilling and unintentional threat to the anti-Catholic reformation and rise of Renaissance thought. His father struggled to educate him. At an early age, Galileo became enamored with a thirst to expand knowledge beyond the limits of the ancients, such as Aristotle's wisdom offered, despite their views that supported traditional church doctrine. Galileo attended a Jesuit monastery to study in a religious environment. Four years later, he informed his father that he wished to be a monk. His father immediately withdrew him from the Jesuit monastery and sent him to Pisa to study medicine in 1581, at the age of 17.

Despite his father's wishes, he chose to study science and within a short time discovered the isochronisms of the pendulum and shortly afterward the concept of hydrostatic balance. These accomplishments earned him a reputation in the field of experimental physics. Galileo joined the faculty of the University of Padua and developed an interest in the astronomical Copernicus theory, discussing the position of Earth's and the sun's rotations and their intergalactic relation to one another. This interest centered on Copernicus's 1543 *Revolutions of the Celestial Orbs*, a work contradicting the Aristotelian belief that Earth was the center of the universe and the sun and all other heavenly bodies revolved around it.

This heresy made Christian dogma suspect. The church had taught that man, a special creation, was the center of the universe. If the Copernican theory were correct, what else in Christianity might God's vicar question? While Galileo's detractors did not accept his interpretations, he remained intellectually safe and secure as long as he did not attempt to spread his ideas publicly. The pope could avoid the issue.

But Galileo recognized the conflict that occurred between science and theology. He corresponded with Johannes Kepler, the eminent German mathematician, who agreed with Galileo's findings but refused to publish these. Kepler's reluctance about political and religious concerns and survival as a price for political correctness appeared warranted. He feared the threats of the Inquisition and the reaction to Copernicus's theories if made public.

Despite the growing storm, Galileo invented a "modern" telescope and in 1609 found confirmation of his and Kepler's Copernican theory. For reaffirmation, he moved his telescope to the garden and observed the Milky Way, the moon and its topography, and Jupiter and its four moons.

Galileo idealistically believed his findings would then be acceptable to society; he apparently did not understand the reasons for the growing tide of opposition to his findings or the stranglehold of theology over science. He became more vocal to his colleagues in Florence where he had recently obtained a teaching position. In 1610, shortly after his appointment, he published his telescopic finds in *The Starry Messenger*. The contents of this work were new for the philosophy of the time. His astronomical findings could have been denied, but since they did not endanger Aristotelian teachings, there was no danger to the church.

His idealism ran afoul when he obtained no support. Educated and cultured people refused to read or discuss the findings publicly. Fear of the church was widespread. He corresponded with Kepler, expressing anguish over lack of concern for the truth. Galileo did not appreciate the power of fear encompassing the nascent Renaissance in Italy. He wrote to his tepid and insincere friend Benedetto Castelio, venting his frustrations and publicly stating that science and knowledge were the most important entities for any society rising from the darkness of the medieval period opposing a church-dominated society. He then carried his struggle to the general public, rather than limiting it to the small

Italian scientist Galileo Galilei defends himself before a court of papal judges in 1633, charged with antiscriptural heresy for championing evidence of the pioneering scientific truth that Earth revolves around the sun. Galileo was convicted and committed to house arrest for the last eight years of life. In 1992, 350 years after his death, the Roman Catholic Church admitted its error and declared Galileo correct. (Oxford Science Archive/StockphotoPro)

cadre of educated people. He hoped the more secular-educated elements of society and the Jesuits would be appreciative and accepting of the discoveries. Why? He wrote compulsively about his discoveries in a style that the "average" reader could comprehend.

He could not be quashed. His lack of awareness and quest for knowledge consumed him. He noticed a lamp swinging while visiting a cathedral and used his pulse to time the movements and found that the swing time was the same for long and short movements. His discovery would help in the development of the clock. He next began a study in displacement and found some minerals were heavier than water, the Archimedean theory of water displacement. While on the University of Pisa staff, Galileo involved himself in a debate over Aristotle's laws of nature. The ancient alleged that heavier objects fell faster than light ones. Galileo challenged this belief by climbing to the top of a tower and throwing several objects of different weights and sizes over the side. They landed at the same time, and another cherished belief of the ancients fell. He then developed a

rudimentary thermometer, a device that would allow a study of temperature variations. The next year he produced a device for raising water from an aquifer, then a military compass for military application used for the accurate aim of cannonballs in war. Not to be outdone, a year later he produced a version of the same implement for land surveying.

Galileo pushed his luck, one that led to religious thought. While at Pisa he gave a presentation on Dante's Inferno. His mathematical skills determined that Lucifer's face was 2,000 arm-lengths and Nimrod's face was as long and wide as Saint Peter's Cone in Rome.

His Pisa University presentation may have inspired the Society of Jesus (Jesuits) to examine his intellectual activities. This society had the responsibility of instructing church creed, philosophy, and now questioning Galileo's scientific findings. Although hostility existed, the church did not feel that he had proven the certainty of his discoveries. No threat existed since the church could allege his findings were hypothetical. In 1611 he visited the Collegium Romanum in Rome and found the Jesuits had replicated his experiments and discoveries. However, Jesuit "scientists" refused to use the telescope, and then Christopher Clavius was instructed to defend the Aristotelian theory. Matters worsened when Galileo and Christoph Scheiner, an important Jesuit, disagreed over sunspots. The session appears to have been intended to protect dogma over science at any cost. This event could be viewed as a warning to Galileo. Later, Tommaso Caccini denounced Galileo from the pulpit for his "heretical" views (Shea 2003).

Further consequences arose in 1613 when Galileo's *The Solar Spots* appeared. It was subsequently attacked by a Dominican friar named Lorini, a professor of ecclesiastical history who vehemently preached that the Copernicus theory violated the holy scriptures that stated Earth was the center of the universe. Sensing the dangers to church theology, Lorini quoted many biblical passages affirming church dogma. Lorini's preaching was so vehement that he was forced to apologize for its intensity. He did not withdraw his concern for the church's position in this moment of danger.

Rather than permit the matter to end, Galileo responded in December 1613 with his *Letter to Castelli*. He averred the truth of the scriptures but argued they must be observed in a figurative sense. He offered an example that defined the "hand of God" as his presence in our lives, not an actual or physical attendance. Galileo cautioned that the Bible was not intended to be a literal interpretation. Further, he cautioned that competing over a correct interpretation of the universe did not attack the holy scriptures. Despite his Solomonic efforts to reconcile the widening chasm between science and theology, all efforts proved futile and the issue became more intense.

Lorini sent a copy of Galileo's *Letter to Castelli* to the Inquisition on February 7, 1615, with a copy of his comments attached. In it, he alleged that a religious in his convent, Saint Mark, believed the letter challenged the scriptures

and Galileo's efforts relegated the scriptures to an intellectual position below natural science. Totally unaware of the firestorm he had created, Galileo expanded his views in a *Letter to the Grand Duchess Christina*. Lorini mollified his actions by informing the Inquisition that Galileo and his supporters were good Christians, but conceited. Lorini's motives were questionable; he opposed those supporting Galileo. Despite his acting as mediator, he still wanted the issue brought before the church.

Lorini's associate, Tommaso Caccini, traveled to Rome to the Holy Office for the purpose of exposing and questioning Galileo's errors and those of his advocates. On March 20, Caccini appeared before the Inquisition and expressed his concern that Galileists' ideas were filling Florence and bringing to question traditional church doctrines and values. Despite his attempts, the efforts of condemnation failed. The attackers had not developed or refined an indictment that could be used by the church. But the church was put on notice of the dangers rising locally and due to the Renaissance in general.

It then attempted to resolve the dispute between Galileo and itself. Cardinal Robert Bellarmine investigated the matter. In a previous instance, the case of Paolo Antonio Foscarini, a Carmelite father, had raised the theory of heliocentrism. In the latter's work, *Lettera . . . sopra l'opinione . . . del Copernico*, he attempted to reconcile Copernicus's views and biblical passages that appeared contradictory. The cardinal stated that Copernicus couldn't be banned but required some editing to make sure the theory was viewed solely as hypothetical.

In a further effort to mollify Galileo, Foscarini sent a copy of his work to Bellarmine who responded to both Galileo and Foscarini. He reiterated that the heliocentric theory was dangerous to the church since it questioned its very foundation. The credibility of those philosophers and scholarly theologians supportive of church dogma believed that any scientific explanation would endanger the church by injuring the holy faith and putting the sacred scriptures into question, a mortal blow to Rome. Ballarmine suggested caution to Rome due to the conflict of science and theology.

Ballarmine, loyal to Rome, offered rationalizations, hoping to calm all and providing a way of resolving the impasse: the perception of the sun's and Earth's movements. He suggested the theories in question be viewed as hypothetical and not the absolute truth with the result that the holy scriptures could be defended. He argued that Galileo's belief in the truth and intellect fell short of absolute evidence and that the scientist had no solid evidence. His views could not be sustained beyond a reasonable doubt.

He learned that truth could not win against church dogma. Given Galileo's intransigence, his appearance before the Inquisition was certain. That institution arose as a self-defensive mechanism intended to funnel free thought and the intellect into the restraint of religious thought. The Inquisition adopted a means of punishment to assist the wayward on their return to the flock. Christianity

had survived splintering long before the reformation, as numerous sects through Europe espoused approaches, which to themselves seemed heretical, such as the Donatists, Cathars, Anabaptists, and the followers of Mithra. However, since the Dark Ages, when Rome became dominant, no variations of thought were acceptable. The Protestant revolution made the situation worse, and Islam gains in the Mediterranean only acerbated the situation.

The church was compelled to act because of Galileo's insistence upon recognition of true science. An appearance before the Inquisition, allegedly a trial, could condemn a free thinker to unimaginable horror in the name of God. This conduct was not new; the church feared deviation from doctrine and authority as far back as its beginning. In 325, Constantine, at the Council of Nicea, empowered the church to bring nonbelievers and those who had fallen away back to the true religion. Lacking any civil authority, the church became the bearer for judging the truth. Free of civil restraints, free thought, dissension, and resistance quickly evolved into a justification for persecution, whether one was guilty or not.

As early as 1231 Pope Gregory IX had become aware of the growing fragmentation within the Catholic Church and resolved to free thought and inquiry through the Inquisition in order to secure the church. In time, the various popes assumed direct control and developed the agencies required to ferret out nonbelievers. Unfortunately, Galileo fell within Inquisitional jurisdiction. The due process rights were nil: the inquisitor brought charges against anyone for unspecified or general charges.

The accused was required to testify against himself after taking an oath on his knees and swearing with a Bible in his hands to tell the truth. He had no right to face his accuser, no right to counsel, sentences could not be appealed, and the accused was examined before two witnesses appointed by the Inquisition. With death or excommunication came confiscation of one's real and personal property, a motive for accusatory action. The testimony upon which the charges were garnered came from criminals, excommunicated persons, and heretics. The accused was given a summary of the charges and took an oath to speak only the truth, the boundaries of which lay in the sacred scriptures.

The means of extracting the "truth" could be torture or fear of torture. The penalties ranged from surrendering your moral values to physical punishment such as submersion in boiling water, loss of limb, or burning and torture in public. The Inquisition could also be lenient: the accused could be compelled to visit church as often as instructed, join pilgrimages, wear a cross of infamy for the rest of his life, or be imprisoned for life.

The accused could not deny the jurisdiction of the Inquisition. The refusal was evidence of guilt. Should the church wish to punish one for his refusal to appear, the suspect was given over to the civil authorities who would torture and probably draw blood. The church justified its conduct by affirming its moral inability to torture one through bloodletting. The civil authorities had no

compunction and used whatever means to draw a confession that could be admitted as evidence of guilt. The secular authority had the responsibility of executing the accused. On occasion, if the accused had died during the proceedings, his body could be burned in public.

The judge, or inquisitor, would institute the proceedings against the suspected heretic. The Inquisition did not address issues of credibility or motives for witnesses. As many members of the community would testify, this was a tactic that strengthened the general fear and provided intellectual control by the Inquisition.

Shortly before Galileo's activities began to concern the church, Pope Paul III, in 1542, established a permanent institution under the supervision of clerics to defend and spread the faith. This church agency, the Congregation for the Doctrine of the Faith, governed the Inquisition's activities. Its governing board included clerics chosen from the Dominican Order.

This body had the jurisdiction to govern Galileo's conduct. On February 19, 1616, the Holy Office decided that Copernicus's theory of an immobile sun and Earth rotating with other planets around it was heresy. Supportive of church dogma, it ruled such thoughts as "foolish, absurd in philosophy, heretical and erroneous in faith." Copernicus's book, *De Revolutionibus Orbium Coelestium*, was immediately placed on the Index, a catalog of forbidden books the church believed would and could harm the faithful and the faith. At the same time, Galileo was admonished for accepting and propagating views not in accord with the church. The inquisitional body, the Qualifiers, issued a report on February 24. It unanimously rendered a verdict that the theory "in regard to theological truth is at least erroneous in faith." The Congregation for the Doctrine of the Faith issued a decree on March 5. Galileo was informed of its actions and the Index condemned the printed documents supportive of the Copernicus theory. This was the body that would finally condemn him in 1633.

Despite this, Galileo met with the pope, who assured him that as long as he remained pope, Galileo would not be surrendered to the Inquisition's jurisdiction. Galileo's actions imply he did not understand the full implications of his approach to science; he apparently ignored the implications of the Inquisition's decision. However, there is some question about the restrictions imposed on him by the Inquisition, for he was allowed to teach in the hypothetical if it did not challenge the sacred sacraments or dogma (McMullin 2005).

Galileo's intellect did not fathom the significance of his actions. He could not convince the church to divest itself from the controversy because the church now made heliocentrism official doctrine. His reactions indicate he didn't understand the power of the church and its cherished doctrines nor did he appreciate the powerlessness of the truth against dogma. Copernicus's *De Revolutibus* was eventually removed from circulation and added to the Index pending revisions.

Galileo continued championing truth apparently without an awareness of the potential intellectual disasters he poised for the church. In 1632, he published

Dialogue Concerning the Two Chief World Systems (the *Dialogue*), intended to offer the public a balanced view of the competing views of Copernicus and the church. When he failed, Copernicus received a prominent position in the discourse. Galileo was ordered to face the Inquisition for trial. However, timing may have been incidental. He had recently published *The Assayer*, a work the church found contrary to the transubstantiation of the Eucharist dogma. Even though Galileo's opponents investigated the charges, no evidence of unorthodoxy was indicated. The Jesuit Order, the right arm of the papacy, was not as lenient as Giovanni de Guevera, a friend of Galileo; it found *The Assayer* as offensive as the *Dialogue*. The Jesuits charged that Galileo's atomism was heretical and contradictory to the Eucharist and protested just as it had over the Copernican doctrine. However, the taint remained and he became more of a target.

The question of what to do with Galileo soon developed an international gloss. Spanish cardinals accused Pope Urban VIII of allowing the Inquisition to be lenient. Urban, in defense, formed a commission to examine the *Dialogue* and to refer its findings to the Inquisition. Urban appointed a commission favorable to his actions relating to Galileo.

Unable to protect him any longer, Galileo was summoned by the inquisitor in Florence to appear before the Holy Office in Rome on October 1, 1632, within a month. The charges were heresy, specifically "for holding as true the false doctrine taught by some that the sun is the center of the world." The "legal" injunction against his condemnation had been sanctioned on February 16, 1616, when he had been instructed to abandon the doctrine in question and not to spread it among those defending it (McMullin 2005).

On April 8, 1633, Galileo was informed that he must stand trial before 10 carefully selected cardinals. He realized, finally, that his only choice was to submit to their decision. Four days later he surrendered to the Holy Office. He faced Firenzuola, the Inquisition's commissary general, who confined Galileo to the Inquisition building until the trial.

When the trial began, Galileo, the penitent, appeared in a white shirt and took the required oaths, and Firenzuola deposed the accused concerning meetings he had had with Cardinal Bellarmine and other church officials in 1616. He couldn't remember all the names of those in attendance 17 years before. He also could not recall the restrictions imposed on him at that meeting. The commissary proposed that Galileo promise not to teach the Copernican theory. Galileo accepted his statement without objection and signed a deposition. He was found guilty and there is no evidence of physical torture. In preparing its sentence, the Inquisition determined his opinions were without foundation and against the holy scriptures. He was instructed to "desist" and not publicly support any of his previous views. He abjured (McMullin 2005).

Weeks later the inquisitors debated the fate of Galileo. He accepted the dictates of the church in the hope he would receive a lighter sentence. Seven

Ptolemy

Ptolemy (90–168) was a mathematician and astronomer in second-century Roman Egypt, best known for his thorough, systematic approach to science, both his own theories and his compilations of current knowledge. His *Almagest* ('Great Book') was the first such treatment of astronomy, and it proposed a mathematically sophisticated model of the universe in order to account for the observable data of his day. In addition to providing a catalog of the visible stars and planets, the *Almagest* described a universe of concentric spheres: each planet (and the sun and moon) existed on its own sphere, rotating around Earth, and beyond the planetary spheres was a sphere of stars in fixed positions. The Ptolemaic Earth does not move—everything else moves around it. Ptolemy did not invent the basic notions of his cosmology, but he had the most data to inform it, drawing on Greek, Babylonian, and Mesopotamian sources. Other astronomers had known their models were faulty because the results predicted by their mathematics didn't match what was observed in the sky. Ptolemy accurately described what could be seen of the world. He remained the dominant figure in astronomy throughout the Middle Ages, for both Europe and the Arabic world.

cardinals demanded his incarceration for life, while a minority suggested he be allowed to modify his *Dialogue* and it could then continue in circulation.

The pope demanded a sentencing and an exact examination of Galileo. When the Inquisition demanded prison for life, Galileo recanted and publicly accepted the Ptolemaic view of the heavens. Galileo concluded his trial with: "I affirm, therefore, on my conscience, that I do not now hold the condemned opinion and have not held it since the decision of authorities. . . . I am here in your hands—do with me what you please" (cited in Linder 2002).

He was sentenced to house arrest for the remainder of his life. As part of the sentence, three members of the Inquisition prepared a seven-page evaluation of the *Dialogue*. They held the book and its author dangerous because they gave life to the Copernican theory, and therefore it was a danger to the life of the church. They had the *Dialogue* burned and banned any of his other works.

A broken man, he was allowed to move to his farm in 1633. He died in 1641, a victim of heresy as the church ordained. He was guilty of a society that opposed free thought and used fear to maintain itself in power. His heresy, convicted by the Inquisition, fell within the parameters of church law for the period.

References and Further Reading

Fantoli, Anibale. *Galileo for Copernicanism and for the Church*, 2nd ed. Notre Dame, IN: Notre Dame University Press, 1996.

Finocchiaro, Maurice. *The Galileo Affair: A Documentary History*. Berkeley: University of California Press, 1989.

Galilei, Galileo. *Dialogue Concerning the Two Chief World Systems*. Translated by Stillman Drake. Westport, CT: Greenwood, 2000 (originally published in 1632).

Linder, Doug. "The Trial of Galileo." 2002. Available at www.law.umkc.edu/faculty/projects/ftrials/galileo/galileoaccount.html (accessed June 14, 2010).

McMullin, Ernan, ed. *The Church and Galileo*. Notre Dame, IN: Notre Dame University Press, 2005.

Shea, William R., and Mariano Artigas. *Galileo in Rome*. New York: Oxford University Press, 2003.

Westfall, Richard S. *Essays on the Trial of Galileo*. Notre Dame, IN: Vatican Observatory, 1989.

11

The Man in the Iron Mask was Count Ercole Antonio Mattioli.

PRO   Justin Corfield
CON   Heather K. Michon

PRO

Following the French Revolution of 1789 and the execution of King Louis XVI on January 21, 1793, the royalist restrictions on the press were lifted, although new ones were applied by some of the incoming republican administrations. Criticism of the arbitrary nature of members of the French royal family abound during the 1790s, and one of the stories that quickly caught the imagination of the public was that after the mob had stormed the Bastille on July 14, 1789, they had found an iron mask there, which, according to popular stories, a prisoner had been forced to wear so that nobody could look on his face. This story spread throughout the world and the prisoner—who was not found at the time of the storming—became known as the Man in the Iron Mask.

However, the story of the Man in the Iron Mask is actually slightly older than this. A prisoner moved from the French state prison at Pignerol in 1687 to the prison at the Isle Sainte-Marguerite had worn a mask, reportedly of steel, and in 1698 a prisoner moved from Isle Sainte-Marguerite to the Bastille in Paris had worn one made from black velvet. The identity of the prisoner was not known for certain, and this led to a large number of theories. The most famous was started by the French philosopher Voltaire, who claimed that the prisoner was a son of Anne of Austria and Cardinal Mazarin, making him an illegitimate older half-brother of King Louis XIV. Because of the nature of the relationship, and, presumably the prisoner looking like either Mazarin or Louis XIV, his face had to be kept from public view. In his book *The Vicomte de Bragelonne*, initially published in 1848–1850 and translated into English as "*The Man in the Iron Mask*," the French writer Alexandré Dumas (1802–1870) transposes the prisoner into being the older twin brother of King Louis XIV. This fanciful theory—encouraged because the prisoner was apparently well treated and shown deference—has led to a number of films based on the book by Dumas. Since then there have been numerous books about trying to trace the identity of the prisoner, some providing detailed explanations of particular theories, others with fanciful suggestions, and others using the genre of novels to explain new ideas.

249

There are a few "facts" known about the Man in the Iron Mask. What is known was that the prisoner was initially held at the fort of Pignerol (modern-day Pinerolo), which was close to the French–Italian border—then a part of France, now a part of Italy. Because of its remote location, the fort was turned into a state prison, operating from the start of French control in 1630. In 1687 the governor of the prison at Pignerol, Benigne d'Auvergne de Saint-Mars, was transferred to become governor of the prison at Isle Sainte-Marguerite, a rocky islet off the coast of Provence. He was notified of his transfer in January and started making arrangements for the move soon afterward. When he did move in May 1687, he took with him one prisoner who was transported in a sedan chair that was covered with an oilcloth. By September it was being reported in Paris that the prisoner in the sedan chair wore a mask made of steel during the transfer. There were also stories that during the prisoner's incarceration at Pignerol, the governor Saint-Mars was responsible for cutting the prisoner's hair and beard when needed, with the prisoner having to wear the mask when anybody else visited his cell. Eleven years later Saint-Mars was appointed as governor of the most infamous French state prison at the Bastille, Paris. He was notified of his appointment in May and made the move in September. On this occasion, he moved, again with one prisoner who was taken in a litter. This time he wore a mask of black velvet.

On arrival at the Bastille, the prisoner was immediately taken to the First Chamber of the Basinière Tower for the first night, and then moved to the Third Chamber of the Bertaudière Tower on the second day. During his time at the Bastille, the prisoner was held in a cell in solitary confinement with Jacques de Rosarges, the second in command of the prison, having the task of personally feeding him. Apparently he dressed in good clothes trimmed with fine lace, and his food was specially cooked for him and served on a silver plate. Lieutenant Etienne du Junca of the Bastille noted that the prisoner did appear on one occasion wearing a "mask of black velvet," presumably to protect his identity. He apparently amused himself by playing the guitar.

The prisoner fell ill in mid-November 1703 and saw a doctor whom he was told was 64 years old. No other details seem to have been given. The man died on the night of November 19–20, 1703, and apparently his head was removed from the body and destroyed, with his body being buried in the grounds of the parish church of Saint-Paul. The burial certificate issued by the priest at Saint-Paul gave the name of the prisoner as "Markhioly," and his age as 45. All the furniture and clothing in the prison cell and any others owned by the prisoner were then destroyed, his silver plate was melted down, and the walls of his apartment were scraped and whitewashed so that there would be no way any inscriptions made by him could have survived. The only object that survived was a loin cloth, which is now on display at the Louvre.

After the French Revolution of 1789, there were attempts by the revolutionary government in France to ascertain the true identity of the Man in the Iron

Mask. By this time the story had caught the imagination of the public with the story highlighting the abuse of French royal power. In 1770 Baron de Heiss sent evidence to the *Journal Encyclopédique* explaining his theory that the prisoner was actually Ercole Antonio Matthioli, an Italian politician. Eight years later Louis Dutens, in *Correspondence Interceptée*, noted that Madame de Pompadour had asked Louis XV about the identity of the Man in the Iron Mask and was told that the person was "a minister of an Italian prince." In 1801, Pierre Roux de Fazaillac, a member of the French legislature, claimed that the Man in the Iron Mask was actually a prisoner Ercole Antonio Matthioli, with the story being an amalgamation of the fate of Matthioli and a valet named Eustace D'Auger; and in 1822 Madame de Campan, who had been a lady-in-waiting to Queen Marie-Antoinette, confirmed the story. In fact a pamphlet published in Leyden, the Netherlands, as early as 1687 had also identified the prisoner as Matthioli.

Ercole Antonio Matthioli was born on December 1, 1640, at Bologna, Italy, and was an ambitious Italian politician who taught law at the University of Bologna. He worked for Charles III, the Duke of Mantua, and was groomed to become his secretary of state. However, before he could get the required promotion, on August 14, 1665, the duke died and his son Charles IV took over. The boy was 13 at the time and as a result, the running of the country was left to his mother, the Dowager Duchess. Matthioli managed to remain a member of the dukedom's senate, with the title of count, but when the duke came of age, he moved to Venice to indulge his passions as a wanton waster. In the power struggle that resulted, Matthioli was forced from office by a monk named Bulgarini, who was also a political secretary and the lover of the Dowager Duchess.

With the duke spending money at such a fast rate, the French decided to use this to their advantage. The Duke of Mantua also held the title of the Marquess of Montferrat, and therefore controlled the strategic fortress of Casale Monferrato. Located on the right bank of the River Po, the castle had been of strategic importance since Roman times. It also had a role in theological history as the place of birth of Ubertino of Casale, a leading Franciscan who plays an important role in Umberto Eco's famous fictional story *The Name of the Rose* (1980). However, it was its military significance that was coveted by France.

The French king Louis XIV wanted to expand his influence in Italy and was keen to control Casale Monferrato, where he could garrison his soldiers and then threaten Turin, the capital of Savoy. However, they had long decided that although the French would win an initial attack on the city, it would lead to conflict not only with Mantua, but also potentially with Savoy and Austria, which were anxious to dominate the region and preferred a weak Savoy to control by the French. The indebtedness of the Duke of Mantua allowed the French to come up with their plan to buy the fort. Indeed in 1659 Cardinal Mazarin managed to buy the Duchies of Nevers and Rethel from the duke's father, and absorb them into France. The French plans were further accentuated when the duke became angry

Illustration from *The Man in the Iron Mask*, based on the story by Alexandré Dumas. (Hopkins, Tighe. *The Man in the Iron Mask*, 1901)

at the Austrians taking over Guastalla in north-central Italy. As a result, King Louis XIV entered into secret negotiations to buy Casale Monferrato. He decided to negotiate through Matthioli, whom he felt could also use it to oust Bulgarini. Matthioli would also be trusted by the duke.

Initially the Duke of Mantua felt he could get the French to pay 1 million livres. However the French decided that the most they would pay would be 300,000 livres, with the agreement made at a masked ball in Venice where the duke was finally forced to agree to accept the French offer. The French would pay half of the money in silver when the agreement was ratified and the remainder when they occupied the city.

Matthioli went to France to negotiate, and King Louis XIV drew up a treaty that Matthioli had to get the duke to sign. Louis XIV was so pleased with the count that he gave him some 200 golden louis and a large diamond. However, soon afterward Isabella Clara of Austria, the Duchess-Regent of Savoy, found out about the secret agreement and contacted the French. By this time the Spanish governor of Milan also knew, and the Spanish offered 600,000 francs for the city. The Duke of Mantua was embarrassed by this and pulled out of the entire deal, leaving the French intensely angry.

It seemed obvious that Matthioli had told both the Savoyards and Spanish, and the Austrians soon found out. Matthioli's exact reasons are not known. It seems that he might have hoped to get one of the other countries to enter a bidding war, and the Spanish certainly did offer more than the French. He may have also been keen to extract some money from the Spanish or another power. The French, embarrassed by this, wanted to get their hands on Matthioli for two reasons. One was that he had a copy of the treaty signed by Louis XIV and awaiting the signature of the duke. The second was quite clearly to get revenge on a man who had betrayed their confidence and made it less likely they would get Casale Monferrato in the short term. However, they did take it in July 1681, with French soldiers moving in on September 30 of that year. They agreed to pay the duke an annual pension of 60,000 livres, appoint him as a general in

the French forces, and make him a part of any French attack into Italy. By that time Matthioli had disappeared.

There is clear evidence that King Louis XIV ordered one of his envoys, Godefroy Comte d'Estrades, to kidnap Matthioli and bring him to France in secrecy. What appears to have transpired was that the French ambassador in Turin invited Matthioli for a hunting expedition outside the city, and during this, he was overcome by 10 or 12 horsemen and taken across the border to Pignerol. It seems the abductors later went into the city to seize his goods and also his valet. In April 1679 Louis stated that "no one should know what becomes of that man," and the French king assured the Duke of Mantua that Matthioli was in their custody and would not be released without the consent of the duke. Matthioli was then held in solitary confinement at the castle of Pignerol (then in the southeast of France, but now in northwestern Italy) and was reported to have gone mad in the following year. Indeed in 1682 a pamphlet titled "La Prudenza trionfante di Casale," written in Italian but printed in the German city of Cologne, detailed the entire controversy involving Matthioli, the treaty, and the count's disappearance.

Matthioli was not the only important visitor held at Pignerol. Nicolas Fouquet, the former superintendent of finances for France, was being held in that prison at the time—indeed when Saint-Mars had become governor of the prison in December 1664, Fouquet was his only prisoner—in the Angle Tower until June 1665 when the tower was damaged and he was transferred to La Pérouse, returning in 1666 to the Angle Tower with two valets, La Rivière and Champagne. Fouquet had been arrested 20 years earlier, in 1661, and after had been taken to Pignerol. Some accounts have suggested that he was the Man in the Iron Mask. Certainly Fouquet was an important man whom Louis XIV wanted held in prison, unable to contact any of his former colleagues, and had the wealth and connections to cause real trouble for the monarchy, being privy to intimate court secrets, but there is a serious problem with this theory. Fouquet died on March 23, 1680, the year before the prisoner wearing the iron mask was transferred to Isle Saint-Marguerite. There is also a surviving French government report, written in 1687, that on May 2, 1679, a new prisoner was taken to the fortress at Pignerol. This 1687 report goes on to state that the man was "the secretary of the Duke of Mantua," undoubtedly Matthioli.

There was also another prisoner at Pignerol—Eustache Dauger, a former valet, possibly the Eustache Dauger involved in high profile poisoning cases in Paris. He was told not to try to communicate with anybody else on the threat of immediate execution—except, of course, Saint-Mars. Other prisoners were also held in Pignerol with Dauger becoming a valet to Fouquet.

In 1681—the year that the French finally took Casale Monferrato—Saint-Mars was transferred to Isle Saint-Marguerite and took with him two prisoners from Pignerol. One of these wore a mask during the transfer. The reason has been presumed to be able to ensure nobody recognized him during his move. Matthioli was certainly at Pignerol before the transfer, and it is unlikely that he

was left there after Saint-Mars moved. This tends to suggest that Matthioli was one of the two prisoners transferred with the governor, and in 1687 there were written accounts that Matthioli was part of the transfer. This leaves the only option being that he was either the Man in the Iron Mask or the man accompanying the Man in the Iron Mask.

If Matthioli was not the Man in the Iron Mask, the question arises as to who that man was. There have been many suggestions. Some reports suggest that Matthioli was kidnapped along with his valet. If these were the two prisoners, it seems extremely unlikely that the valet would wear a mask and Matthioli not.

Saint-Mars, in a report in 1688, noted that some people were claiming that the prisoner was Richard Cromwell, the son of Oliver Cromwell. However, by this time Richard Cromwell had returned to England where he was living quietly, perhaps happy that rumors still placed him in France. It was also suggested that he was James, Duke of Monmouth, who had been defeated in 1685 at the Battle of Sedgemoor by his uncle James II, king of England. Some felt that James II would not have executed his nephew, but the main problem with this theory is that the Man in the Iron Mask makes his first appearance in 1681, when Monmouth was still in the limelight in England and the United Provinces (the Netherlands). It has also been suggested that since Nicolas Fouquet had died, perhaps the Man in the Iron Mask was a valet of Fouquet. This is curious because most of the arguments against the Man in the Iron Mask being Matthioli presumed that he was not well-known enough to need to have his face hidden. If Matthioli was not infamous enough, then a valet of Fouquet could hardly be more infamous.

In 1687 the next move of the Man in the Iron Mask took place. On this occasion, Saint-Mars moved from Isle Saint-Marguerite to the Bastille in Paris, having been promoted as governor to the main French royal prison. This time he took with him only one prisoner who wore a black velvet mask. Soon after this transfer a pamphlet was published in the Dutch city of Leyden stating that the prisoner in this transfer was Matthioli. It is not known how widely the pamphlet was circulated.

A letter to Saint-Mars from the secretary of state in 1697 cautioned the governor not to "explain to anyone what it is your long time prisoner did." This has led to some people feeling this is enough evidence to show that the Man in the Iron Mask was not Matthioli, as everybody knew what the Mantuan count had done. However, there is the intriguing possibility that Matthioli might also have been engaged in a few more complex political games than has hitherto emerged.

Soon after this move, the political situation in Mantua had changed considerably. With Charles IV, Duke of Mantua, becoming a firm French ally, the duke had joined with Louis XIV in the War of Spanish Succession and was declared a traitor by the Austrian Emperor Joseph I in 1701. Two years later Charles IV visited Paris and during that visit, the Man in the Iron Mask died at the Bastille prison and was buried nearby. Some have seen this as a possible coincidence. The most convincing explanation that Matthioli, often described in contemporary

---

The Sun King and the Iron Mask

Rumors about the parentage of Louis XIV were rampant long before they were immortalized in literature. Many stories of different possible identities for the Man in the Iron Mask have made their rounds through France and the world, but most of them, with the possible exception of Ercole Antonio Matthioli, had to do with attempts to discredit Louis XIV's legitimacy as king of France. The enmity between Louis XIII and his wife, Anne of Austria, was well known. They had been married for 23 years before their son, and the only heir to the French throne, was born. Rumors flew that the Duc de Beaufort was the younger Louis's actual father, a fact picked up on by Alexandré Dumas in his *Three Musketeers* novels. It was even believed by some that it was Beaufort in the iron mask for all of those years. In another version of the iron mask myth, the son-in-law of Anne's doctor, Marc de Jarrigue de La Morelhie, was imprisoned in the mask because he accidentally stumbled upon evidence showing Louis's dubious nonroyal parentage.

French documents as Marthioly, was the man who died in 1703 was that the burial register lists the dead person as "Markhioly," and his age as 45. However, when a doctor had seen him soon before he died, he gave his age as 64, roughly corresponding to the age of Matthioli (who would actually have been 63 at the time). Some have suggested that because so much energy went into preventing people from ascertaining his identity, the name in the burial register must also be false, but it is just as likely that the name was genuine. There was a fanciful theory that the name used was actually a contraction of "Marquis Ali," referring to Nabo the Moor, a Moorish dwarf valet to Louis XIV's wife, Marie-Therese, who seems to have had an affair with her and then vanished from the court soon after the birth of a child believed to have been fathered by Nabo. However, it seems unlikely that Nabo the Moor would have had to have his face disguised by an iron mask—there were very few Moorish dwarves around at the time, and the mask, iron or velvet, would hardly hide who he was for long.

With the War of Spanish Succession continuing, in April 1706 the French commander in the south, Marshal de Vendôme, seized the military initiative and drove back the Savoyards and their Austrian allies and started the siege of Turin, the capital of Savoy. It seemed that Duke Charles IV's alliance with the French was about to pay off. Eugene of Savoy took to the field against the French and by August had reached Turin. On September 7, at the battle of Turin, Eugene led an unexpected attack on a badly deployed wing of the French army and destroyed the French forces who, over the next two months, withdrew from Italy, retreating across the Alps. It was one of the greatest campaigns of the war, and in 1708 all the lands of the Duke of Mantua were formally seized and handed to the House of Savoy. Charles IV, the last of House of Gonzaga, died in July of that year at Padua. He had married twice but had no children.

The French ambitions in Italy lead one back to the question as to why the French decided to keep Matthioli in such secrecy that he needed to become the Man in the Iron Mask. The traditional view has been that the iron mask was used to conceal the identity of the prisoner because his face was well known. This led Dumas to write his account that the prisoner was a twin brother of Louis XIV (or indeed Louis XIV himself, after a "swap" at court). The face of the king of France was on every coin in the realm, and there were paintings of him around the country. Certainly few people would be able to recognize the face of Matthioli, and with it being well known that he was held a prisoner of the French, the purpose might have been to conceal exactly where Matthioli was being held. The punishment of using the mask might also have served as a warning to other people from double-crossing the French, and the French were always eager to get their hands on the original of the treaty that Matthioli had hidden before his capture. Two other possibilities have also emerged, both also pointing to Matthioli as the Man in the Iron Mask.

It has been suggested that the reason why Matthioli might have been held with a mask was not to conceal his identity forever. The fact that a prisoner was held in a mask quickly became well known, and the notoriety would cause people to ask questions. Although the prisoner might have initially been Matthioli, when the state wanted to hide a more senior prisoner, Matthioli (in the mask) could be killed and another prisoner could be substituted at a later stage. Most people would continue to believe that it was still Matthioli and would be none the wiser. This conspiratorial theory means that Matthioli held in the iron mask was essentially the means to disguise future arrests should they take place.

There is also another theory that arose to why the mask was used. Part of this can be explained by the fact that there was probably no iron mask. It is now believed that the Man in the Iron Mask actually wore a velvet mask rather than an iron one and also had only worn it when there was some chance of him being seen by other people. There is clearly the possibility that the wearing of the mask might not have been insisted on by the authorities but chosen by the prisoner to conceal his identity. Matthioli had been negotiating in Venice where masks were worn at dances, and indeed the masked ball was used as a place where French agents could meet the Duke of Mantua. Alternatively it could have been revenge by a vengeful French king who, having been tricked by Matthioli, thought a masked ball might have sought to punish him in an unusual manner. Having incarcerated Matthioli with orders for his face to be hidden, the king might not have concerned himself with the fate of a man who had tried to trick him. Courtiers might then have been afraid to mention the matter in the same way as, in the 1970s and 1980s, the family of Moroccan Defense Minister Mohammed Oufkir was held in prison for 15 years essentially because nobody felt themselves able to bring up the fate of the family of a traitor to King Hassan II.

There also remains the possibility that Matthioli might have worn the mask at his own request. Both instances where the mask is known to have been worn

were when going from one prison to another, and Matthioli might have been keen to avoid being recognized by the Savoyards or others whom he betrayed, or he might have worn it to attract attention for himself. Although these are possible, they in no way detract from Matthioli being the Man in the Iron Mask.

Many historians have accepted the identification of Matthioli as the Man in the Iron Mask. Marius Topin, in his book *L'homme au masque de fer*, published in 1870; and Franz Funck-Brentano, in his book, originally published in 1894, both believed that Matthioli was the Man in the Iron Mask after detailed investigations. Indeed Topin suggested that it was likely that a mask—as used for masked balls—was among the possessions of Matthioli when he was seized; there is evidence that the French also seized his baggage at the same time as they kidnapped him, for no other reason than to seize the copy of the proposed treaty Matthioli had been negotiating. Therefore, he might have worn the mask during the transfer for his own reasons, in much the same way that modern prisoners often try to hide their faces from press and television cameras. It might have been that Matthioli had worn the mask to achieve what he had not in his own life. By trying to disguise himself in such an obvious manner, and one which would certainly draw considerable attention to himself, he was trying to do just that—draw attention to himself in the hope that some of his former political allies in Mantua or elsewhere might come to his aid. The last word had best be left to Prince Michael of Greece (1983) who, in his sympathetic biography of Louis XIV, wrote that Matthioli was "probably the dullest and least romantic, but the most plausible" of the candidates for the Man in the Iron Mask.

References and Further Reading

Funck-Brentano, Franz. *Legends of the Bastille*. Ingram, TN: Fredonia Books, 2003.

Michael of Greece, Prince. *Louis XIV: The Other Side of the Sun*. London: Orbis, 1983.

Noone, John. *The Man Behind the Iron Mask*. Stroud, UK: Alan Sutton, 1988.

Topin, Marius. *L'homme au masque de fer*. Paris: Didier and Cie, 1870.

von Keler, Theodore. *The Mystery of the Iron Mask*. Girard, KS: Haldeman-Julius, 1923.

CON

L'Homme au masque de fer. The Man in the Iron Mask. For three centuries, his identity has remained one of the great unsolved mysteries of history. The rich and powerful of the 18th century made a virtual parlor game of guessing his name and his crime, while commoners turned him into a symbol of royal tyranny.

Historians have devoted countless hours in dusty archives looking for the answer, while novelists, playwrights, and filmmakers have turned him into the tragic centerpiece of their novels, plays, and motion pictures. Yet for all the time spent and all the ink spilled, little more is known about the prisoner today than on the moment he appeared on the historical stage.

The Prisoner

That moment was the afternoon of Thursday, September 18, 1698. Lieutenant Etienne du Junca, a prison official at the Bastille in Paris, noted the event in his journal. At three o'clock, the newly appointed governor of the prison, Bénigne d'Auvergne de Saint-Mars, arrived with his luggage, his staff, and, "in a litter, a longtime prisoner, whom he had in custody at Pignerol, and whom he kept always masked, and whose name has not been given to me, or recorded." On Saint-Mars's directions, de Junca had prepared "the third room in the Bertaudière tower." There, the unknown prisoner was isolated. "The prisoner," du Junca noted, "is to be served and cared for by M de Rosarges," second in command to Governor Saint-Mars.

He made only a few appearances in du Junca's journal before ten o'clock on the night of November 19, 1703, when the "unknown prisoner who has worn a black velvet mask since his arrival here in 1698" died after a single day's illness. Saint-Mars had him buried in the graveyard at nearby Saint Paul's, recording his name in the parish records as "Marchialy" and his age as about 45 years.

The Legend of the Masked Man

An unknown person dying a lonely prison death and consigned to a poorly marked grave in a parish cemetery should have been quickly forgotten. Yet just eight years after his death, the story of the masked prisoner had reached as far as the royal household. Princess Palantine was the sister-in-law of the French king, a prodigious letter-writer with an ear for gossip. On October 15, 1711, she wrote to her aunt Sophie, *heiress presumptive* to the British throne, that "a man has lived for long years in the Bastille, masked, and masked he died. Two musketeers were at his side to kill him if he unmasked. He ate and slept in his mask. There must, no doubt, be some reason for this, as other wise he was very well treated, well lodged, and given all that he wanted. He received Holy Communion in his mask; he was very devout and was perpetually reading. No one has ever been able to find out who he was." But just one week later, on October 22, the princess was back at her writing desk with the answer. "I have just learned who was the masked man who died in the Bastille. His wearing a mask was not due to cruelty. He was an English lord who had been mixed up in the affair of the Duke of Berwick (natural son of James II) against King William.

He died there so that the King might never know of him." Case closed as far as the princess was concerned, although a quick check of the records shows her conclusion was a false one.

In 1717, a 23-year-old budding writer and philosopher then known by his birth name of François-Marie Arouet was sent to the Bastille, falsely accused of penning a poem that had scandalized the French court. During his 11-month stay at the prison, he heard the story of the masked prisoner from guards and inmates who had seen or heard of the man during his five years in the Bertaudière Tower. Thirty-four years later, Arouet, now known by his pseudonym Voltaire, memorialized the story in his *Le Siècle de Louis XIV* (The Age of Louis XIV). The black velvet of du Junca's journals was replaced by the less soothing image of an iron hood permanently riveted over the prisoner's head, with a "movable lower jaw held in place by iron springs that made it possible to eat wearing it."

In his 1751 history, Voltaire only hinted at the prisoner's presumed identity. The man, he said, had been treated with extreme courtesy, with staff removing their hats when they entered his cell and politely refusing offers to sit with him while he ate. The governor, Saint-Mars, brought his meals and linens to him. He ate off a silver plate, and his cell was nicely furnished and decorated. In short, the prisoner was treated like someone of high social rank—like royalty.

Voltaire waited another 20 years before divulging whom he believed the man to be: in his 1771 *Questions sur l'encyclopedie*, he stated that the prisoner was none other than the bastard brother of Louis XIV, the illegitimate offspring of Anne of Austria and Cardinal Mazarin. Unlike many that would follow, the story at least carries a kernel of authenticity. Louis XIV ascended to the throne of France in May 1643, just before his fifth birthday. His mother acted as his regent during his youth and adolescence, and she had chosen Cardinal Mazarin to run her regency government. They were probably lovers, and some even believe they secretly married. However, it would have been very difficult to hide a pregnancy in the gossipy French court, particularly given the fact that neither Anne nor Mazarin was very popular. Louis XIV took power when Mazarin finally died in 1661 and ruled France for the next 54 years.

Voltaire's was just one iteration of the prisoner's life that put him within reach of a throne. The most famous variation on the theme was the one immortalized by Alexandré Dumas in his 1850 novel *The Man in the Iron Mask*: that the prisoner was the older, identical twin brother of Louis XIV, hidden away to prevent a power struggle for the throne. Other versions made him the illegitimate son of Anne of Austria, this time with the English Duke of Buckingham, or the illegitimate son of Louis XIV himself. Others thought he might be the illegitimate son—or daughter—of the Queen Mother. (The "man" as a woman appears off and on throughout mask literature.) Still others thought he might have been the Duke of Monmouth or some other illegitimate son of England's King Charles II, or the French playwright Molière. In 1801, the utterly fantastic rumor

Alexandré Dumas

More than anyone else, Alexandré Dumas has shaped the modern myth of the Man in the Iron Mask. Although the historical accuracy of the myth has never been taken seriously by historians, the idea that the king of France's older brother could be forced to live out his days inside a prison wearing an iron mask to hide his true identity was originally put forth by Voltaire, later written about by Michel de Cubieres, and finally popularized in the serialized novel written by Dumas during 1848–1850. The novel, actually one volume of a larger work titled *Le Vicomte de Bragellone*, is a sequel to his earlier and most famous work, *The Three Musketeers*. He wrote many other works, including his popular *The Count of Monte Cristo*. But one of his less popular works proved to be his most daring: in *Georges: The Planter of the Isle of France*, Dumas takes on the issue of mixed races (Dumas himself was part African in his ancestry) and European colonialism.

began to spread that the prisoner, again Louis XIV's hidden brother, had fathered a child while imprisoned, and the expectant mother had returned to her native Corsica to give birth to their bastard child: an ancestor of the future emperor, Napoleon Bonaparte.

These stories, and others like them, were everywhere by the end of the 19th century. In 1901, Tighe Hopkins noted that there were about 40,000 treatises on the prisoner in the famed library of the British Museum, each with their own twist on the tale. Reading at the impossible rate of one a day, he pointed out, it would take a man over 100 years to get through them all (Hopkins 1901). In the 1930s, French historian Maurice Duvivier estimated that at least 1,000 people had been named as candidates (Noone 2003).

While ferreting out the identity of the Man in the Iron Mask was a pleasant diversion for armchair historians, for the common people of France, it had become more than an interesting tale to tell over the fire on the cold winter's night. To many, it was yet another symbol of royalty run amuck. That a man could be incarcerated for a lifetime on the whim of a king, stripped of his name, his identity obliterated behind a mask while he lived on for year after year, summed up much of what was wrong with the system. When the mobs finally stormed the Bastille in July 1789, over eight decades after the prisoner's death, rumors quickly spread that the rioters had found the skeleton of a man with an iron mask still bolted on his skull deep in the bowels of the hated fortress before they burned it to the ground.

Two Candidates Emerge

By the early 20th century, researchers had dispensed with virtually all of the prisoner-as-royalty stories. It was an understandable presumption: the man must have been masked because he resembled someone who was famous, and in an

era before photographs or even widespread distribution of paintings or drawings, there were very few famous faces outside the king and the royal family. From there, the many stories that emerged seem like a logical next step. The second-tier candidates were those who might not resemble somebody famous, but might have been famous in their own right. This was a much smaller field, with even Voltaire noting that "there did not disappear from Europe any personage of note" when the prisoner was believed to have first been sent to jail in the 1660s or 1670s (cited in Hopkins 1901).

Drawing on one of the few facts that seemed constant throughout all the various tales of intrigue—that the prisoner had been under the perpetual care of Bènigne d'Auvergne de Saint-Mars during his long incarceration—researchers found only two prisoners who might possibly fit that profile: Antonio Ercole Mattioli (or Matthioli) and a man registered as "Eustache Dauger."

The Problem with Mattioli

At first glance, Mattioli seems to be the stronger case. A politician in the court of Charles IV, Duke of Mantua, a region in what is now northern Italy, he had become involved in negotiations over the control of the fortress at Casale Monferrato, overlooking the strategic Po River Valley not far from the French border. Louis XVI offered 100,000 crowns for Casale, and Mattioli brokered the secret deal between the duke and the king so effectively that Louis gave him several lavish rewards. But just as the French were about to set out for Casale, Mattioli double-crossed both parties, telling the governments of Venice, Savoy, Spain, and Austria about the deal, probably in hopes of getting greater rewards for his information. Louis was forced to withdraw from Casale and issued an order for Mattioli's arrest, adding an ominous note to the warrant: *Il faudra que personne ne sçache ce que cet homme sera devenu*—"See that no one knows what becomes of this man." Seized by French agents, he was deposited at the prison at Pignerol under the pseudonym "Lestang" and kept under the harshest conditions, provided with only the bare necessities for survival. He was later transferred to the Fort Royal prison on the isle of Saint-Margaruite (Hopkins 1901).

Members of the royal family who came after Louis XIV seemed to think Mattioli was the man in question. Louis XV is said to have told Madame de Pompadour that the prisoner was a "minister of an Italian prince," while Louis XVI told Marie Antoinette he was a political figure from Mantua.

There are, however, several problems with Mattioli as the Man in the Iron Mask. The king's comment is likely not as ominous as it first seems: to capture Mattioli, his agents had to seize him from Mantua, which could anger the duke or the people of the region. Mattioli's imprisonment was never a secret; accounts of it appeared in the newspapers as early as 1682. His treachery was well known. More important, he did not follow Saint-Mars when the governor was transferred to Les

Entrance leading to the dungeon of the "The Man in the Iron Mask" at the Castle If, before his transfer to Sainte-Marguerite. (Branger/Roger Viollet/Getty Images)

Exiles in 1781 and wasn't under his control again until Saint-Mars was sent to Saint-Margaruite in 1687. And Mattioli/Lestang vanishes from prison records by the end of 1694, a good indication that he died sometime that year and making it impossible for him to be the masked figure who arrived in Paris four years later.

The Search for Eustache Dauger

This left Eustache Dauger. It was long believed to be a pseudonym—many state prisoners were held under false names—but on the assumption that it was a *real* name, historian Maurice Duvivier (1932) spent several years combing the archives of the Bibliotheque Nationale in Paris looking for it or any of its likely variations. There he found Eustache Dauger de Cavoye.

Cavoye was born into the minor gentry of the Picardy region in 1637, but grew up at the royal court. His father, Francois Dauger de Cavoye, was a musketeer who rose to the rank of captain of Cardinal Richelieu's guards. He married a young lady of the court, and they had six sons before his death in battle in 1641. As a child, Eustache was a playmate of King Louis XIV, who was about his age.

All six Cavoye sons joined the military, with four ultimately dying in battle. Eustache himself had participated in more than a half-dozen military campaigns by his 21st birthday, but in most respects, he was clearly the black sheep of the family. In 1659, at the age of 22, he was found to have participated in a "black

mass" held at the Castle de Roissy on Good Friday. Black masses were Satanic, or at least profane, corruptions of the Catholic Mass, and this particular event created a huge scandal, with several military officers losing their positions as a result.

Cavoye escaped punishment due to his connections within the court—his brother Louis was one of the king's chief officials—but it was just the first in a string of catastrophes. His mother all but disowned him in her 1664 will, naming the younger Louis as head of the family. In 1665 Cavoye was forced to sell his army commission after killing a man outside Saint-Germain in Paris following a drunken altercation. Finally, he was caught up in a "poison scare" that swept Paris in 1668 or 1669. Two characters named Le Sage and Abbé Mariette were charged with witchcraft after evidence emerged that they were dispensing love potions and poisons and performing black masses. Some of the highest ranking ladies of the court, including Madame de Montspan, the king's mistress, were rumored to be implicated, as was Cavoye, who may have assisted in delivering poisons and setting up black masses.

With his brother Louis in the Bastille after an ill-advised love affair that had led to a duel, there was no one to protect him. He was supposedly arrested at Dunkirk while trying to escape to England and sent to the prison at Pignerol.

Duvivier's theory was strong: Eustache Dauger de Cavoye's crimes were shocking in the context of the times, the type of thing to land a person in prison for life. His childhood relationship with the king would have made it difficult for Louis XIV to order his execution. At the same time, he was not someone who was so recognizable that his disappearance would be notable.

However, later research showed that while Eustace Dauger de Cavoye was imprisoned sometime around 1669, it was most likely by his own family, and he was sent not to Pignerol, but to Saint-Lazare in Paris, where he died around 1683 (Duvivier 1932).

The Valet

Scottish author Andrew Lang had another candidate. Governor Saint-Mars was told by the French secretary of war that the prisoner he needed to keep under such close guard was "only a valet, and does not need much furniture." Lang found records in the British archives that made reference to a servant to a Frenchman living in England in 1669 named Roux de Marsilly. The valet was known only as "Martin" (Lang 1903: 9).

Roux de Marsilly was a Huguenot, a French Calvinist. From their beginnings during the reformation in the early 16th century, the Huguenots had struggled to gain power, or at least religious freedom, within the Catholic kingdom. This fight had cost thousands of lives, both on the battlefield and in periodic civilian massacres; tens of thousands of Huguenots left France for new homes as far away as South Africa and the Americas. The conflict eased with the 1598

Edict of Nantes, which granted the Huguenots religious freedom and civil equality within the realm. But the self-styled "Protector of Catholicism" Louis XIV had steadily reversed the rights granted under the edict and would revoke it altogether with the Edict of Fontainebleau in 1685.

Marsilly was one of the many Huguenots working to bring about an alliance with Protestant countries to try win their French brethren greater freedom, either through diplomacy or force. Records in the British archives indicate that he was corresponding with many Protestant sympathizers in England and on the continent. He often used his valet Martin as a messenger.

When Marsilly was arrested by the French in Switzerland in the spring of 1669, the rumor spread that he was part of a widespread plot to kill King Louis XIV. He was taken to the Bastille, where he refused to divulge anything about his activities in England. On June 22, 1669, he was publicly tortured and executed. Just before the guards came to take him to the courtyard, he slashed himself with a shard of broken mirror. Officials simply cauterized the wound with a hot iron and continued on. He was stripped naked and staked out on a wooden frame. Government officials made several attempts to get him to confess to an assassination plot, but Masilly denied he knew anything about it, telling his interrogators "hee never had any evill intention against the Person of the King," according to an English onlooker. He died about two hours later (Lang 1903).

After Marsilly's execution, French officials were keen to talk to Martin and find out what he know of his master's intrigues. On July 1, 1669, Charles Colbert de Croissy, ambassador to England's Court of Saint James, wrote Paris that "Monsieur Joly has spoken to the man Martin and has really persuaded him that, by going to France and telling him all that he knows about Roux, he will play the part of a lad of honor and a good subject." There is some evidence that Martin was not so easily persuaded, apparently telling his interrogators that by returning to France, he realized people would believe he had knowledge about Marsilly and he "would be kept in prison to make him divulge what he did not know."

Sometime between July 2 and 19, Martin vanished from the records, and on July 19, the valet "Eustache Dauger" emerged. That day, Louis XIV's war minister, the Marquis de Louvois, informed Saint-Mars that the safekeeping of his new prisoner, then being held at Dunkirk, was "of the last importance to his service." Dauger was to be held in confinement, with no contact with the outside world. His meals were to be served to him by Saint-Mars himself. "You must never, under any pretenses, listen to what he may wish to tell you. You must threaten him with death if he speaks one word except about his actual needs." Saint-Mars promptly replied that he had told Dauger he would run him through with his sword if the prisoner spoke so much as a single word.

Despite this rude introduction, in 1672, Saint-Mars asked that the prisoner be assigned as a valet to the Duc de Lauzun, a high-ranking inmate destined to spend almost a decade at Pignerol. In his letter, Saint-Mars said that Dauger did

not want to be released and "would never ask to be set free." Permission was denied, and Louvois ordered he have no contact with Lauzun or other state prisoners. The only real hint of discord came in 1680, as Dauger was tossed into the prison's dungeon. "Let me know how Dauger can possibly have done what you tell me," Louvois cryptically asked Saint-Mars, "and how he got the necessary drugs, as I cannot suppose you supplied them."

Generally, Dauger seems to have been little trouble. He "takes things easily, resigned to the will of God and the King," said Saint-Mars in the late 1780s. He spent most of his time in solitary confinement, spoken to only by Saint-Mars and his lieutenant, a confessor, and a doctor who was allowed to examine the prisoner in Saint-Mars's presence. When Saint-Mars was transferred to command of the prison at Saint-Margurite near Cannes, he transported Dauger in a sedan chair, his body covered by a piece of oilcloth—a 12-day ordeal that very nearly killed him. On the trip to Paris, he was carried on a litter, eating his meals and sharing a room with Saint-Mars as their group stopped each night at country inns.

No matter which prison he was kept at, "l'ancien prisonnier" (the old prisoner) was rarely seen by prison staff. His cells were comfortable, but not lavish, and despite later myths, he seems to have eaten off the same plain plates as all the other prisoners.

The lingering mystery of what Martin/Dauger had done to warrant such careful confinement was never spelled out in any of the correspondence. Saint-Mars guarded Dauger for over 30 years and claimed not to know anything of his identity or his crime. Andrew Lang came to the conclusion that his decades-long punishment had been "the mere automatic result of the 'red tape' of the old French absolute monarchy," a victim of bureaucracy, who was forced to endure years of isolation by a government that neglected to rescind an order long after the prisoner had ceased to be a threat. (Lang 1903: 5)

Not only was the Man in the Iron Mask not masked with iron, he may not have been masked throughout his imprisonment. It is not mentioned in any known correspondence between 1669 and 1698, nor was it part of any known directive by the French government. The first time the mask appears is on the road to Paris in 1698. Apparently, it remained in place for the last five years of Dauger's life.

Conclusion

It seems clear from the evidence that the valet Martin was the Man in the Iron Mask—a servant who lived a much stranger life than most. Nothing else is known of him, not his first name or his date of birth, not where he came from, or his religion. His thoughts or feelings about his unique fate are a mystery. What few possessions he carried through his 34 years in jail were destroyed, and his cell sanded and repainted in the days following his death, as went the standard

procedure of the time. But in the words of author Tighe Hopkins, he "arises and steals from that graveyard of St. Paul's this ghost that shrouds his face, intent upon an odd revenge, the torment and insoluble conundrum of historian, fabulist, novelist, dramatist, essayist and gossip—the Sphinx of French history: the Man in the Iron Mask." (Hopkins 1901: 24)

References and Further Reading

Barnes, Arthur Stapylton. *The Man of the Mask: A Study in the By-Ways of History*. London: Smith, Elder, 1908.

Davenport, R. A. *History of the Bastille and of Its Principal Captives*. London: Thomas Tegg, 1839.

Dover, George Agar-Ellis. *The True History of the State Prisoner, Commonly Called the Iron Mask*. London: J. Murray, 1826.

Duvivier, Maurice. *Le masque de fer*. Paris: Armand Colin, 1932.

Francq, H. G. *The File of the Man behind the Mask*. Winnipeg, Manitoba: Francq, 1984.

Furneaux, Rupert. *The Man Behind the Mask. The Real Story of the "Ancient Prisoner." [An Attempt to Identify the "Man in the Iron Mask" as Eustache Ogier de Cavoye. with Plates.]* London: Cassell, 1954.

Hopkins, Tighe. *The Man in the Iron Mask*. Leipzig: Bernhard Tauchnitz, 1901.

Lang, Andrew. *The Valet's Tragedy: And Other Studies*. London: Longmans, Green, 1903.

Macdonald, Roger. *The Man in the Iron Mask*. Cambridge, MA: Da Capo, 2005.

Noone, John. *The Man behind the Iron Mask*. London: Palgrave Macmillan, 2003.

Topin, Marius Jean François. *The Man with the Iron Mask*. London: Smith, Elder, 1870.

Von Keler, Theodore. *The Mystery of the Iron Mask*. Girard, KS: Haldeman-Julius, 1923.

Williamson, Hugh Ross. *Who Was the Man in the Iron Mask?: And Other Historical Mysteries*. London: Penguin, 2002.

Wilson, Colin. *The Mammoth Encyclopedia of the Unsolved*. New York: Carroll and Graf, 2000.

12

Prince Louis Charles (Louis XVII), also known as the "Lost Dauphin," survived captivity during the French Revolution and was allowed to escape in 1795.

PRO   John Lee
CON Lorri Brown

PRO

In the mysteries of history, Louis XVII's story is the most tenacious of them all and the most evocative of the many sympathetic emotions that are aroused by tragedies because of the youthfulness of the victim of the mystery. His name is surrounded by the storm of the French Revolution at its most sadistic and inhumane state when, driven by fear of foreign royal intervention, they executed all known threats, including their own members of the democratic revolution, which started in 1789. They created various instruments to root out enemies of the state. As a result, Louis XVII was imprisoned and suffered a horrible death because of the callousness of his prisoners. However, many still believe and claim that this never happened, or it is clouded with mystery so one cannot decide the issue. After his death, there arose many persons who claimed to be the surviving prisoner who had escaped from the prison, and this section will explore these claims.

The history of Louis XVII (March 27, 1785–June 8, 1795) is a horrible tale of intrigue, deception, and revolution that has been kept alive because of the various people who have claimed to be the lost prince of the French throne. The French Revolution (1789–1793) culminated in the overthrow of the French monarchy by the middle and lower social classes of France because Louis XVI could not manage the national debt and finances to help his people. It enacted a constitution and an assembly that guaranteed the rights of men. However, the later part of the revolution became more radical and denounced all royal power, and the people decided to overthrow the monarchy. This was called the Reign of Terror because all of the enemies to the democratic state (royalists, dissenters, and political parties) were rounded up and executed by the guillotine. As a result, the nation became so antimonarchial that it laid the foundation for what would follow with the demise of the royal family.

The incident with Louis XVII happened at the height of the French Revolution when the Reign of Terror was beginning for France. Robespierre, the leader

267

of the Reign of Terror, had instituted several government committees to protect the internal security of the government and its leaders from foreign and domestic conspiracies to overthrow the new government. This new French revolutionary government was still very insecure because they had planned to dethrone Louis XVI and his wife, Marie Antoinette. They were afraid that the European monarchies would invade France to ensure the safety of the king and queen. The revolutionary government began the overthrow of the king by capturing Louis XVI and imprisoning Marie Antoinette in the Temple Prison, a medieval fortress in Paris, which was built by the Knights Templar in the 12th century and was made into a prison by the revolutionary government. Antoinette, a Habsburg princess, had three children by Louis XVI: Louis Joseph Marie Therese, and Louis Charles (who became Louis XVII). They imprisoned Louis XVII, the king's son, in the same tower with his mother, and he became another victim of the French Revolution hatred of royalty. However, Louis XVII also had his supporters in France and around the European continent, who had proclaimed him to be the king of France at the death of his father. After Louis XVII, they would continue to support various people who claimed to be Louis XVII, the "Lost Dauphin" (heir to the throne of France). As a result, there have been over 300 claimants for the title of Louis XVII.

Louis Charles or Louis XVII was born on March 27, 1785, to Louis XVI and Marie Antoinette during the French Revolution. He was called Louis Charles, Duke of Normandy (1789–1791). Then he was named Louis Charles (1791–1793). He was finally named *Fils de France* (son of France) on his birthdates. His position as heir to throne was the reason he gained the title of Dauphin, in French meaning heir to the throne. During his first seven years of his life he had a nanny named Agathe de Ramband, Berceus de Enfant. Madame de Rambaud took care of him during those seven years and was in charge of his discipline and health needs. Then Louis Elisabeth took care of him for the next three years. He grew up in the trappings of royalty until the night when the republican government of the French Revolution overthrew his royal family. The Committee of General Security was a police agency charged with overseeing the local police in Paris. It was to supervise internal enemies and anyone suspected of treason. The committee sent the family to the Tuilleries and then to the Temple Prison. The Dauphin was separated from his father on December 2, 1793, during his father's trial. On July 3, 1793, commissioners or representatives of the Committee of General Security arrived to divide the royal heir from his mother and sister, Marie Theresa. It is said that Marie Antoinette refused to leave her son and protected his body with hers, but the representatives threatened to kill her and her daughter if she did not release him. She saw the hopelessness of the situation and let Louis go with his captors. Louis cried and begged for his mother to rescue him, but he was dragged into solitary confinement one floor beneath Marie Theresa. At that time, the royalists, or the people loyal to the

throne, had declared Louis to be the next king of France, but the republicans wanted to stop the possibility of him becoming king. So the little king was only 10 years old when the forces of the republican government abducted him.

There are several accounts of what happened to Louis XVII in the Temple Prison. Many accounts describe how he was subjected to sadism, human indignities, and other horrible experiences in the name of the justice of the Reign of Terror, but there is no proof or recorded evidence that this happened. Louis was kept in a room with Antoine Simon and Marie Jeanne, who were cobblers and revolutionists. According to stories, Antoine Simon treated Louis XVII with contempt and abuse. The Dauphin's head was shaved on January 5, 1794. He was made to testify against his mother in a trial and accuse her of incest and other charges that caused his mother more emotional troubles. In addition, Simon made the son of France drink alcohol and eat in excess. He also made Louis curse like a lower class citizen as well as wear the clothes of the French peasants. They also made him sing the *Marseillaise*, the national anthem of the French republic. They made him sleep with a prostitute from whom he contracted venereal disease. Louis was also threatened with the guillotine to exact obedience from him. Another story was that the Simons told Louis that his parents had disinherited him and did not like him. He was also called "Capet" as an insult to his royalty, because the kings of France were never called by their surnames, but rather by their royal names. It was used as a term of criticism and to make him feel like a commoner. These stories were never confirmed but were made by the royalists after the French Revolution. Other reports have him being well treated by Marie Jeanne and well fed by this kind woman. However, some believe that the revolutionists may have covered up this story.

In 1794 the Simons departed from the prison. The head of the Committee of General Security, Jacques Foche, visited Louis XVII. Later, Jean Jacques Christopher Laurent (1770–1787), the representative of Josephine de Beuharnais, future wife of Napoleon I, saw Louis XVII and released him for a walk and to breathe in clean air. He said that he was in good health and saw no bruises. In March 1795, three government commissioners, J. B. Haund de la Men, J. B. L. Matthener, and J. Devensche, visited him and saw nothing wrong with him either. This may be proof that the reported atrocities were false reports made by the royalists.

On May 3, 1795, he was reported to be seriously ill. Some revolutionary accounts state that he was found in the fetal position, full of puss and unable or unwilling to speak because of the abuse that he had suffered. Another republican said that he was sensitive to sound, had diarrhea, and was tubercular. He died on June 8 of tuberculosis, but some said he had a skin problem called scrofula. His body was immediately examined by one of the government's physicians and he was pronounced dead. The physician also took the opportunity to extract the child's heart and place it in a bag, which he took with him when he left the house. It was an old French tradition to separate the heart from the body of the French royalty.

The body was not buried, and the king's son was pronounced dead to the French people by the revolutionary government. The physician kept the heart for many years in a glass jar and then he lost it at a bridge in Paris. The physician recovered the jar many years later and it was kept in a church in Paris until recent times.

Marie Therese, his sister, was not allowed to visit the Dauphin or to say prayers for him even though she was in the floor above him. Some false claimants have considered this evidence that the Dauphin could have escaped or that the revolutionary government was attempting to cover up a murder. Many rumors started circulating that the French government had murdered him. As frequent as these rumors were, other rumors circulated that he was alive and had escaped the terrible fate of his parents. It is also reported that Louis XVIII sent out many imposters to deflect other people from finding the Dauphin, because he was a competitor to the throne. In addition, he also prevented Marie Therese from investigating rumors and imposters for the same reasons. It was a sad ending for the Bourbon family. As the years past, thousands of articles and books were written on the disappearance and possible survival of the Dauphin.

Pretenders to the Throne

One of the first claimants to the thrown was a tailor's son, Jean-Marie Hervagault (September 20, 1781–May 8, 1812), who said he was removed from the temple in a linen basket. He was the illegitimate son of Duc de Valentois, Prince of Monaco. Other claimants boasted of being removed by means of concealment inside a rocking horse. Hervagault convinced many people that he was 13 when he was really 15 at the time of imprisonment. When the bones of the king were dug up, it was said they looked like they were those of a 15-year-old. Many aristocrats, clergy, and Joseph Fouchs, Duke of Otranto, supported his claim. He also gathered a court at Chalons sur Marne in France that respected him as king of France. However, the claimant's royal life was stopped when he was tried for swindling and thrown into jail with another pretender, Mathuran Bruneau, who was famous for pretending to be other nobles. The royal family did not back his claim, and his honesty is in question because of his criminal charges.

There is another pretender, Ethelbert Louis Hector Alfred, Baron de Richemont, who wrote a dark account of his own escape from the tower in 1832. Richemont (Ethelbert Louis Hector Alfred) claimed that he was Louis in Paris in 1828. In 1842 he was arrested and thrown in jail for 12 years for a minor crime. He escaped and left the country until his return in 1830. He died at Gleize on August 1853 and the name Louis Charles de France was placed on his tombstone until the French government removed it.

The best imaginative story was that of the English famous naturalist, John James Audubon, who was truly suspected to be Louis Charles. Audubon was the naturalist who wrote the *History of Birds in America*. He was an adopted child and

was about the same age as the Lost Dauphin at his adoption. Audubon also wrote a letter to his wife that said, "patient, silent, bashful, and yet powerful of physique and of mind, dressed as a common man, I walk the streets! I bow! I ask permission to do this or that! I . . . who should command all!" (quoted in Tyler 1937: 57). This story is fantastic in its emotional power, but it is not very believable.

There are other stories of the Dauphin being transported to America to hide with the Native Americans. This fantastic story is a claim by Onwarenhiiaki, a Native American who claimed to be the Dauphin after being transported to America by his cousin the Duke of Provence to live with the Iroquois. Another false claimant was a man named Eleazar Williams. He was born and raised in Wisconsin, but he claimed that he was the Dauphin. He said he had amnesia until the age of 13 and had no memory of his childhood. He lived with the Native Americans in New York and became a missionary to them. But in 1850, he claimed that he was the Lost Dauphin and attempted to claim the riches that were associated with the Dauphin. Williams claimed that the Dauphin was concealed in the fourth story of the tower and a wooden figure had replaced his body in his coffin. They also had a deaf mute who would imitate him sitting in the window. He also claimed that A. Barras, who was a friend of Josephine de Beuharnais, future empress of France with Napoleon, decided to save the Dauphin for future reign in France. He said that he concealed the Dauphin in the tower and the Dauphin left in a coffin and was taken from the cemetery. Some said that he looked like the Comte de Provence and he was brought to the United States by the ex-valet of Louis XVII. The city of Green Bay, Wisconsin, erected a statue in honor of Eleazar that claimed that he had similar scars to that of the Dauphin. However, Eleazar Williams was discredited and the royal family did not claim him as their son.

The Lost Dauphin in Wisconsin?

Among the many people claiming to be the Lost Dauphin during the early 19th century was Eleazar Williams, one of the early settlers in Wisconsin Territory during the 1820s. Williams was, at various times, an unpopular religious leader, a spy, an Indian negotiator, and a frontier settler, quite a résumé for one who, by the late 1830s, was claiming to be the legitimate heir to the French throne. Williams did have a physical resemblance to members of the Bourbon family that ruled France until the Revolution, which may have spurred him to make his claims. In 1841 Prince de Joinville, the youngest son of King Louis Phillip of France, made a tour of the United States, which gave Williams an idea. He arranged to meet the prince because he had been a Native American missionary. However, Williams would go on to use the meeting to claim that Joinville had actually come to see him, knowing his "true" identity. Williams claimed that the prince offered him extravagant wealth if he would only renounce his claim to France's throne. Of course, the prince denied all aspects of the story.

Louis XVII was the titular king of France from 1793 until 1795. He was the son of King Louis XVI and Marie Antoinette and spent his childhood in prison during the early years of the French Revolution. His treatment there damaged his health, and he died at the age of 10. (Library of Congress)

His true history was that he is a descendant of a Mohawk Native American and a white woman who had been kidnapped by the Mohawks at seven years old. He grew up with the Mohawks, but left as a teenager to become an Episcopal minister and a pioneer to Green Bay, Wisconsin. Also, Eleazar wrote a book about his claim that he titled *The Lost Prince*, which brought him fame for several years. He did claim until his death that he was Louis Charles, but there was no evidence to support this claim. His head was later exhumed and DNA testing confirmed that he was Native American and thus was not Louis Charles.

There is another story that stated an imposter had approached Marie Theresa, who was walking with her brother-in-law, the Duc de Bury, in a garden near Versailles. The imposter called out to her, "Sister!" Marie Theresa reportedly replied, "Go away! Go away! It is you who is destroying our family." The imposter ran away from the princess, but it left Marie Therese in hysterics, because of the unsettled nature of her brother's death.

Throughout history there have been between 40 to 300 claimants or pretenders to the throne. These included a person name Lazarre William of New York, another who claimed to be the Dauphin living among Native Americans. The many false claims made illustrate the gripping and emotional opportunity that pretending to be the Dauphin had for his pretenders. There was a drive in those who attempted to take advantage of the royal family and their riches for their own gain. All of these attempts by pretenders have cast some doubt of ever knowing the true fate of the true Dauphin.

Possibilities for Louis XVII

In spite of these false claimants, there is one who came the closest to being the true Dauphin. This claimant provided evidence that may sufficiently identify him as the Dauphin. He provided photographic evidence as to his similarity to

those of the royal family, and he may be the physical generation of the Bourbon family and possibly the Dauphin himself.

Karl Wilhelm Naundorff (1785?–August 10, 1845) was a German clock- and watchmaker who claimed to be the Lost Dauphin. He lived in Spandau in Berlin, Germany, where he became a citizen. In 1822 he was in Brandenburg but was jailed for three years for arson. In two memoirs written in 1827, he wrote that there was a substitute in place of the French king who had escaped in the 1820s to Germany. In 1833 he went to Paris and claimed that he was the son of Louis XVI. Naundorff went to Paris where he spoke to members of the royal family and told them much about their private life with him. He convinced the wet nurse Agatha de Rambaud that he was the Lost Dauphin. People of the court like Etienne de Joly, Louis XVI's minister of justice, and Jean Bremond, the king's personal secretary, identified him as the Lost Dauphin. They showed a picture of Naundorff to Marie Theresa, his sister, but she did not recognize him. In 1836, Naundorff sued Marie Theresa for property that belonged to the Dauphin, but Louis Phillippe, the Bourbon king at this time, arrested him and deported him (Cadbury 2003: 100). On January 1, 1860, he thought he would be restored to the throne, but he died before that dream came true. His relatives put an epitaph of his grave stone that read "Here lays King of France, Louis XVII." His relatives have attempted to press his claims in the French courts during the 19th and 20th centuries. Rene Charles de Bourbon, his grandson, has attempted the claim in the French court, but it was rejected. His family has attempted to put forth his claims on the Internet, where pictures of Naundorff, his father, and his brother are placed beside the pictures of Louis XVII's family. There is a close resemblance especially in the shape of the nose and the lips, which they claim identifies Naundorff as the Lost Dauphin.

There were serious rumors that he was still alive at the time of 1819. There was also proof that he was still among the diplomats who signed the French treaty of 1819 that ended the revolutionary government. Eckert, a diplomat, claimed that many wise men refused to give up the claim that he survived. He also stated that many of these claims suited the royalist, who wanted to elevate the Dauphin to the throne, so the French government never pursued the claims. He also stated that it was odd that the members of the family never wore black or attended his funeral, as if they knew he was still alive. These rumors still continue to haunt the memory of Louis XVII.

The legend of Louis Poiret of the Seychelles may be another story indicating that the young Dauphin was still alive after his imprisonment in the tower. His story is that he was taken by a man named Poiret to the Seychelles near Madagascar on the Indian Ocean where he lived his days in Poivre away from his enemies. During the early 19th century, he supposedly worked with the inhabitants on the cotton industry. Then in 1822, he moved to Mahe, another island, and claimed that he was the son of Louis XVI and Marie Antoinette. In 1857, he

was on his deathbed and told his relatives that he was the Dauphin. There are those in the Seychelles who claim they are descendants of Louis XVII.

These three cases are the closest to being true claimants to the throne of France, because of the physical similarities and the intimate details that the people provided to prove their authenticity.

The DNA Evidence

In 2000 the leading historian of the Lost Dauphin, Philippe Delorme wanted to have the heart that had been preserved tested for DNA evidence to resolve the controversy. He even considered double checking the result through two tests conducted by Jean Jacques Cassimen, professor of genetics at Belgium's Louvain University, and Ernst Brinkman, professor of Genetics at Germany's University of Munster. The testing of the Dauphin's heart is also an interesting contention against the claimants. They took the desiccated heart of the Dauphin and compared it to preserved hair from Marie Antoinette and Marie Theresa. They compared the mitochondrial DNA of each specimen and discovered that there was a 75 percent match between the DNA of the samples (Cadbury 2003: 135). However, Delorme stated that the heart could have belonged to a relative of Marie, but Delorme still feels that there is sufficient evidence that the heart is that of the Dauphin and considers the case closed. It convinced enough historians and the French government that they buried the heart between the coffins of Marie Antoinette and Louis XVI. His descendants attended the funeral and proceedings that interred the heart with his parents.

Some skeptics have cast doubt on this evident matching of the DNA by claiming that you cannot determine inheritance through the mother, the heart may have been replaced during the lost period of time between the discovery of the heart and now, and the samples may be too old. The history of the heart is also problematic, because after Pellatin stole the heart, he pickled it in alcohol. Then his student stole the heart and dropped it in a river, but returned it to the doctor after a period of remorse. His wife sent it to the archbishop of Paris where it stayed until the palace was attacked in the Revolution of 1830. The doctor retrieved it. In 1840, the restoration began and the heart was sent to Spain to be with the Bourbon branch of the monarchy. The crystal vase was smashed during the restoration, but the doctor's son retrieved it. It was returned to Paris and finally placed in a crystal vase at the Royal Crypt at Saint Denis where it rested until 1999. In 2000, when Delorme tested the heart, is was described as being as hard as wood.

Critiques of the DNA Test

As a result of this research, Phillippe Boiry (author of a book on Naundorff) remains skeptical that any real DNA could have been obtained from the hardened

surface of the heart, especially since it has endured many exposures to the elements, in particular being dropped by the student into the river. Phillippe de Boiry (2000) claims that there might have been two hearts involved in the escape from the tower. One of the hearts may have belonged to Louis Joseph Xavier François, the first son of Louis XVI, who had died in 1789. In addition to this historian, many relatives of the three major claimants remain doubtful about the DNA testing. Naundorff's relatives requested his grave to be reopened to do more testing.

One last claimant actually wrote a response to an e-mail claiming that his ancestor, related to Count Esterhazy, a friend of Marie Antoinette, had taken the child to Hungary. He gave the jailers a large amount of gold to release the prisoner. He was taken to Hungary where Louis XVII lived the remainder of his days unable to forgive the French people for their atrocities toward his family. Thus, the odyssey of the Lost Dauphin continues even up to today.

Conclusion

The tragic story of the Dauphin and his pretenders has continued for many years since that horrendous time at the tower. It has shown how much interest and passion there is in the mystery of the Lost Dauphin. Many claimants have attempted to persuade the royal family and prove that they are their lost relatives either for money or family ties. Also, although there was conclusive evidence through DNA testing, many people still deny the physical evidence, because the heart that was tested was in poor condition and had endured many trials. As a result, it cast doubts on the identity of the heart as well as whether the Dauphin was still alive. Even to this day, there are still claimants to the throne of the king of France that may continue into the future as long as people are interested in history's mysteries and the romantic era of royalty.

References and Further Reading

Bordonove, Georges. *Louis XVII et l'enigme du Temple*. Paris: Pygmaillian, 1995.

Cadbury, Deborah. *The Lost King of France: How DNA Solved the Mystery of the Murdered Son*. New York: St. Martin's Press, 2003.

De Boiry, Phillippe. *Le Dossier: Louis XVII, une affaire de couer*. Paris: F-X. de Guibert, 2000.

Hardin, John. *Louis XVI, The Silent King*. New Haven, CT: Yale University Press, 1992.

"Heart of Boy Gets Royal Funeral in France." *China Daily* (June 3, 2004). Available at http://www.chinadaily.com.cn/english/doc/2004-06/03/content_336250 .htm (accessed June 15, 2010).

Platon, Mircea. "The Reeducation of the Dauphin." *Convorbiri Literare* (May 2004). Available online at http://convorbiri-literare.dntis.ro/PLATONmai4.html (accessed August 16, 2010).

Royale, Madam, Duchess of Angoulême. *Private Memoirs of What Passed in the Temple, From the Imprisonment of the Royal Family to the Death of the Dauphin*. London: John Murray, 1823.

Stone, R. J. "Death in the Temple." *Quadrant* (June 2005).

Tyler, Alice Jaynes. *I Who Should Command All*. New Haven, CT: Framamat, 1937.

Watelle, Michelle. *Louis XVII ou le secret de roi*. Boucherville, Québec: Editions de Mortagne, 1990.

CON

In June 2004, following an elaborate state funeral, a crystal urn containing a small, petrified human heart was officially laid to rest in the royal crypt of Saint Denis Cathedral in Paris. The heart, believed by both scientists and historians, belongs to Louis XVII, otherwise known as the "Lost Dauphin." For 209 years, controversies raged about whether Louis escaped during the French Revolution or died in Temple Prison. For all intent and purposes, the official state funeral for Louis XVII should have laid to rest any question about his death. However, this is one historical controversy that continues to be debated.

The tale of the Lost Dauphin is one of the best-known controversies in world history. It began during the French Revolution, in the Temple Prison of Paris, where young Louis Charles was held with his family for six years before his death. The true fate of Louis XVII, know to historians, scientists, and royalists as the "Lost Dauphin," has been long scrutinized. Even before his untimely death, there was speculation that the little king had escaped. Some claimed that sympathetic jailers, who spirited him away to Germany, England, or even the United States, secretly rescued Louis from his prison. Others believe the boy king, who was but 10 years old when he died, did die in Temple Prison and a physician secretly took his heart during the autopsy. Indeed, it is the heart of Louis XVII that is at the center of the controversy. For over 100 years there was no way to determine whether Louis XVII had died or escaped. Over time the true facts of the Lost Dauphin melded with legend and lore. However, thanks to modern science, historians have been able to rule out imposters claiming to be Louis XVII and to positively identify the heart in Saint Denis. DNA tests were first used in 1950 to disprove a German clockmakers claim to the French throne. In 1999 further DNA tests were conducted to prove once and for all that the heart belonged to the Lost Dauphin, Louis XVII (Cadbury 2003).

According to the official history of France, Louis XVII, the titular king of France, died of tuberculosis in the Temple Prison in Paris on June 8, 1795.

However, few people accepted tuberculosis as the cause of death. Some believed neglect and harsh treatment by his jailer, Antoine Simon, killed the king while others whispered of poison and murder. Some declared he wasn't dead at all, but had been smuggled out of the prison and taken abroad. The rumors surrounding the boy king's death have continued far beyond the French Revolution. Theories of Louis XVII escaping France to Germany, England, or America have persisted for over 200 years. During the 19th century over 100 men claimed to be the lost king of France, ranging from the probable to the impossible. Some of the better-known claimants included John James Audubon (although he never made a public claim) and a Wisconsin missionary by the name of Eleazer Williams. Williams's claim was never taken seriously, though there is a Lost Dauphin Road, Lost Dauphin Park, and Lost Louie's Restaurant in the town of Lawrence, Wisconsin, in his honor. Anthropological studies on the exhumed skull of Eleazer Williams indicate that he was of Native American descent and therefore rule out the possibility of his being the Lost Dauphin. Another claimant from the Seychelle Islands is included in the "official" history of the islands. The 19th-century author Mark Twain parodied these far-fetched claims in *The Adventures of Huckleberry Finn*. The most plausible candidate claiming to be the Lost Dauphin was a German clockmaker by the name of Karl Wilhelm Naundorff. Naundorff was able to convince the childhood nurse of Louis XVII that he was indeed Louis XVII. He was officially recognized as a Louis XVII by the Netherlands government and was allowed to take the surname Bourbon. However, genetics testing on a bone removed from the remains of Naundorff in the 1950s proved he did not share the same DNA as other relatives of the Bourbon family. Even after the Naundorff case was laid to rest by the DNA evidence, conspiracy theories that Louis XVII escaped and had spent his life abroad continued to circulate. It wasn't until 2004, 209 years after the Dauphin was declared dead, that his heart was finally laid to rest at the Basilique Saint Denis in Paris.

The story of the Lost Dauphin stretches back to his early childhood. Born on Easter Sunday, March 27, 1785, the second son of Louis XVI and Marie Antoinette, Louis Charles was not originally the Dauphin of France. The title of Dauphin belonged to his older brother, Louis Joseph, who would inherit the thrown. Right from the beginning, the little prince of France was surrounded by controversy. Whispers through the lavish court at Versailles and beyond to the city of Paris claimed his mother, who was highly unpopular with the French people, had been unfaithful to her husband. Critics, political enemies, and even relatives claimed that a Swedish nobleman, Axel, Count Von Fersen, was really the new prince's father (Fersen 1971). To his credit, Louis XVI ignored all rumors and innuendos that his newest son was not really his and rejoiced in his growing family. The Dauphin, Louis Joseph, had never been in strong health and by the summer of 1789, his health began to fail rapidly. He died on June 5, from a

combination of tuberculosis, rickets, and pneumonia. Four-year-old Louis Charles then became the Dauphin of France and heir to the French throne.

The state of France's economy at the time of Louis Charles's birth would play a central role in his short life. For years prior to his birth, French peasants and merchants were unfairly taxed by the government. Thanks to the unequal voting of the Estates General, any attempt at economic reforms were cut short by the nobility, who were lightly taxed. Almost none among the taxpaying public in France had political power, and the very few who made up the nobility, or the Second Estate, effectively exercised veto power over any tax reforms. Taxes on the commoners were kept high, and this, combined with shortages of bread, set the stage for rebellion. Added to this was the newly established United States of America (whom Louis XVI had aided through raised taxes). The French peasants were inspired by the Americans' success in winning freedom from Great Britain and taxation without representation.

While economic conditions rapidly deteriorated during the 1780s, the royal family lived an insular life at Versailles, seven miles outside of Paris. Louis Charles and his older sister Marie Therese, along with their parents, lived a life of unprecedented opulence at Versailles attended by an army of servants and courtiers. While the majority of people in Paris and the rest of France lived in dire poverty, the French Crown spent great sums of money on the intricate rituals of court life. It is estimated that to serve the royal children a simple bowl of soup cost 5,000 livres, because of the cost of the retainers required to actually serve it. At the center of the people's growing discontent were the king and queen. The queen especially received a great deal of ugly attention, thanks to propaganda published by political enemies. In these scandal sheets, called *libelles*, the lives of the royal family were common fodder, with numerous accusations that the queen had trysts with many other nobles who were at court. The fact that Louis XVI did not consummate the marriage for seven years only added to the thought that the queen was being unfaithful. Even after the birth of two children, the rumors persisted. The queen's supposed dalliances led to her portrayal as an aloof monarch who, like Nero, "fiddled" while Paris starved. Marie Antoinette was shown as a promiscuous, frivolous queen, secretly working for her Austrian relations rather than for France. Despite the fact that Marie Antoinette visited poor houses and hospitals, trimmed down her own lifestyle greatly, and gave to various charities, her public image did not improve.

During the summer of 1789, France was in the throes of heated political turmoil. Thunderstorms had destroyed the previous year's crops, making wheat scarce and causing massive bread shortages throughout France. As the summer dragged on, hot and marked with severe storms, discontent among the people grew by leaps and bounds. There were loud rumbles of revolution in the air. In an effort to bolster his own claim to the throne, King Louis XVI's vile cousin, the Duc de Orleans, had bought up large quantities of flour to increase the

effects of the shortage. He also hoped to further increase discontent against Louis and Marie Antoinette. All of these events left an uneasy atmosphere in Paris. Added to the problems was the fact that Louis and Marie Antoinette were hardly seen by the public. There were accusations that the king and queen were plotting against the people and were held up in secret meetings. What the public did not know was that the king and queen were with their dying son, away from Versailles at Meudon.

Following the storming of the Bastille on July 14, 1789, through the end of the summer into autumn, the French aristocracy began fleeing abroad to other parts of Europe. However, the king and queen stayed in Versailles with their children. Louis XVI continued to try and work with revolutionary forces, to reach a compromise, but was unable to. On October 5, 1789, mobs descended on Versailles. Crowds screamed obscenities at the king and queen, and in the end the royal family was forced to leave Versailles. Louis Charles would never see his childhood home again. Following the royal family's removal from Versailles Palace, they were housed together at Tulieres Palace in Paris, where the revolutionaries could keep a close eye on them. Louis XVI and Marie Antoinette, along with their children, spent four years under house arrest. Their situation did not improve after a failed attempt at escape in June 1791, when they were captured near the German border in the town of Varennes. The royal family remained intact in their palace prison until Louis XVI's trial and execution on January 21, 1793. Little Louis Charles, not quite eight years old, was now considered by royalists to be Louis XVII, king of France. Several European countries also formally recognized the new king, including England, Portugal, Austria, Russia, and even the United States.

In the dead of night on July 3, 1793, guards burst into the royal family's rooms and demanded to take Louis away from his mother and sister. Marie Antoinette refused and pleaded for nearly an hour before finally relinquishing her son. According the Marie Therese, the guards "threatened the lives of both him and me and my mother's maternal tenderness at length forced her to this sacrifice." Louis Charles would never see his mother again.

Little is known of Louis's incarceration at the Temple Prison. According to her memoirs, for many nights following the little king's removal, Marie Therese and her mother could hear his cries and screams, as he was tortured and beat by the guards. While most of the writings about the little king's upkeep were destroyed during the French Revolution, it is generally acknowledged that his captors were cruel, to say the least.

At first Louis was placed in the care of a cobbler by the name of Antoine Simon by the Committee of General Society. Simon and his wife, Marie Jeanne, were instructed by the government to remove any signs of "arrogance and royalty" from both the king and his sister. Simon was to bring the boy up as a "good *sans culotte*," in reference to the name given to the patriots of the French

Temple Prison in Paris where the "Lost Dauphin" (Prince Louis Charles) was held during the French Revolution. (Apic/Getty Images)

Revolution. The government hoped to coerce damning evidence out of the boy that they could use against his mother at her upcoming trial. There is much debate as to the extent of the cruelty inflicted on the little king by the Simons. It is generally acknowledged that the Dauphin received clean clothes, decent meals, and was allowed to play with toys retrieved from royal storehouses. He was even allowed to play with the laundry woman's daughter every two weeks, when she came to fetch the linen for washing. However, there is no doubt that the Simons were not fit guardians for Louis Charles. There were no school lessons. Instead they taught the little king to sing lewd songs, to use vulgar language, and eat and drink in excess. Simon even asked officials what they wanted done with the boy. "Citizens, what have you decided about the wolf cub? He has been taught to be insolent, but I shall know how to tame him. Hard luck if he dies because of it! I will not answer for that. After all, what do you want done with him? To deport him? To kill him? To poison him?" (Cadbury 2003: 103). In October 1793, the Simons had managed to coerce those in charge to make charges of incest against his mother.

While the French Revolution raged outside of the Temple Prison, the "Orphans of the Temple," as Louis Charles and his sister Marie Therese were known, were locked away, unseen by the public. The siblings were not even aware of the execution of their mother on October 16, 1793. Madame Simon fell ill in January 1794, and she and her husband promptly left the Temple Prison. From this point on, the records of Louis XVII imprisonment were largely destroyed, shrouding the rest of his story in mystery.

The sudden departure of the Simons gave rise to rumors that the Simons had smuggled the king out of the prison. At this time Louis, declared to be in good health, was given over to the care of the guards. With no particular guardian, Louis Charles was not seen or spoken to for as long six months. While Marie Therese was kept in relative comfort, accounts of the little king's upkeep indicate he was placed in a tiny cell and fed a poor diet. His cell was not cleaned nor was he allowed to bathe. Food was pushed through a small barred *guichet* in the lower half of the door. There was a small window in the cell, but with walls that were several feet thick, it was hard for light to penetrate the gloom of the prison. Those

who believe that the king escaped Temple Prison often cite the inconsistency of the harsh treatment given to both Louis and Marie Therese. Of course, Marie Therese could never inherit the French throne, as it passed only through the male line. Therefore, it was little Louis Charles who represented a very real threat to the fledgling republic, and this explains his cruel treatment. As long as he lived, there would always be a chance the French monarchy could regain power.

Under the deplorable living conditions, Louis XVII's health quickly began to deteriorate. Before he was removed from his mother and sister, Louis had grown into an attractive boy, with fair curls and big blue eyes. However, during his two years of isolation from his family, vermin, bugs, fleas, and scabies covered his skin, rendering him quite unrecognizable. He did not grow properly, with his legs and arm disproportionately long and with rounded shoulders. His wrists, elbows, and knees swelled from infection. This only helped to add credence to stories of his escape. It would have been easy to replace the king with another boy similar in age and appearance. Not until July 1794 was a new guardian appointed to the king, Jean Jacques Christopher Laruent. At this point the king was bathed, clothed, and his room cleaned. In December 1794 government officials finally visited the boy. By this time Louis Charles refused to speak. This spurned new rumors that the king had been replaced by a deaf mute.

The royal siblings became useful pawns for the new French government. Outside pressure from the rest of Europe began to demand reports of both Louis XVII and Marie Therese, as part of the formal recognition of the new French government. Within France people began to question what was happening to the brother and sister. The government was quick to squash any rumors of ill treatment of either Louis Charles or Marie Therese. Journalists who dared to criticize the care of Louis and his sister were promptly thrown into jail. In March 1795 a new guardian was appointed to the king by the name of Etienne Lasne. On May 6, 1795, Pierre-Joseph Desault was allowed in to examine the king. The doctor reported the atrocious conditions under which the Lost Dauphin was being held. Less than a month later Desault was dead. The cause was a mysterious illness, but there were mutterings of poison and murder. Two more doctors were called in to care for the king. Jean-Baptiste-Eugenie Dumangin and Philippe-Jean Pelletan tried unsuccessfully to treat all of Louis's ills. At about two in the afternoon on June 8, 1795, 10-year-old King Louis XVII died. Two of Louis Charles's jailers witnessed the cadaver, and the following day government officials came to see the body as well. However, none of these officials ever saw the Dauphin in person, and Marie Therese, the most plausible candidate to confirm her brother's death, was not allowed to view the corpse. The twisted little body was buried on June 10, 1795, at the cemetery of Saint Margurite in an unmarked grave.

News of the king's death was not released until four days later. Following the announcement of Louis XVII's death, many people who had assisted in the torture and harm of the little boy offered up weak apologies. Some even

suffered mental breakdowns. Madame Tison, who worked with her husband at Temple Prison, was one who suffered a breakdown. She insisted that she had helped the king escape. Madame Simon, who had participated in tortures of the king, claimed the king had visited her in her hospital room. "My little prince is not dead," she claimed. These confessions, which continued to spill forth for many years following the king's death, kept public interest in a possible escape theory alive. Many people questioned why the Dauphin had received such inhumane treatment while his sister was left relatively alone. It did not make sense, and increased further speculation that the Dauphin had escaped.

While performing the autopsy, Pelletan, the attending physician, secretly removed the heart of Louis XVII and wrapped it in a handkerchief. Removing the king's heart was in keeping with royal tradition. Pelletan then stored the heart in distilled wine alcohol. In 1815, following the restoration of the Bourbon kings, Pelletan offered the heart to Louis XIII (uncle to Louis XVII). The new king of France refused the gift, however. Pelletan then gave the heart to Monsieur de Quelen, the archbishop of Paris, who kept it hidden in his private library. Following the plundering of the archbishop's library in 1830, Pelletan's son retrieved the heart of Louis XVII from among the rubble and stored it in a crystal urn, where it remains today. Pelletan's son died in 1879 and the heart passed through a series of hands, including the Spanish branch of the Bourbon family. In 1975 the great-granddaughters of Don Carlos de Bourbon returned the heart to Paris to the Duc de Bauffremont, president of the memorial of Saint Denis. Even then the heart was not recognized as that of Louis XVII, but rather a heart of a child lost during the French Revolution. It would take 25 more years to prove it was the heart of the Lost Dauphin.

Following the restoration of the French monarchy in 1814 and throughout the rest of the 19th century, men from all over Europe and even the Americas claimed to be the Lost Dauphin. Karl Wilhelm Naundorff, a German clockmaker, was able to convince several people close to the royal family that he was indeed King Louis XVII. He was interviewed by Madame de Rambourd, who had acted as Louis XVII's maid. Naundorff told her of several incidents at Versailles that few outside of the royal family would know. He also bore a remarkable resemblance to Louis XVII, sharing the same scar on his lip, certain moles, and vaccination marks. Despite all the evidence laid forth, Marie Therese never met with Naundorff. Naundorff made such a fuss about his claim to the throne that he was eventually thrown out of France. He went into exile in England and eventually settled in the Netherlands. He persisted until his death in 1845 that he was indeed the true Louis XVII. The Dutch government officially recognized Naundorff as Louis XVII and allowed him to use the surname Bourbon, and his descendants still carry the Bourbon name. His death certificate even reads "Louis Charles de Bourbon, aged 60, son of Louis XVI and Marie Antoinette." Naundorff's widow went so far as to sue Marie Therese for her late husband's "inheritance," which

A Claimant Makes His Case

One of the many people claiming to be the Lost Dauphin, Augustus Meves, tried to prove his claims by publishing his memoirs as *The Authentic Historical Memoirs of Louis Charles, Prince-Royal, Dauphin of France* in 1868. The following is his claim:

> The Dauphin's escape from the Tower of the Temple is no longer a mysterious problem. The truth that has hitherto been enveloped in a labyrinth of obscurities is now substantially removed. The demise of the son of Louis XVI and Marie Antoinette in the Tower of the Temple was but a Republican assertion and not an historical fact, which has benefited the ruling sovereigns of France, since the sacrifice of Louis XVI, to acquiesce in, namely, Napoleon Bonaparte, Louis XVIII, Charles X., the Dukes of Angouleme and Berri (the sons of Charles X.), Marie Thérèse (the daughter of Louis XVI married to the Duke of Angouleme, consequently in the interest of the house of Charles X), le Cornte de Chambord, the acknowledged legitimate sovereign of France (son of the Duke of Berri), Louis Philippe, the Orleans Family, and, lastly, Louis Napoleon—all have had, and still have, an interest in recognising the authenticity of the Republican announcement, that the son of Louis XVI died in the Temple. Nevertheless, with all this apparent antagonistic array of political influence, with its attendant legions, to annihilate truth, such will be impotent and ineffectual, for sooner or later it will manifest itself; and though it may be lulled to sleep for a while, time will eventually dispel the illusions, that chicanery and artifice have invented.

Source: Augustus Meves. *The Authentic Historical Memoirs of Louis Charles, Prince-Royal, Dauphin of France*. London: William Ridgway, 1868, vi.

was never granted. The Naundorff case has been one of the strongest in the argument that Louis XVII escaped Temple Prison during the revolution.

There have been two major tests conducted to determine if Louis Charles died in Temple Prison or escaped his captors. The first test was conducted in 1950. The right humorous bone from Naundorff's grave was tested against strands of Marie Antoinette's hair (taken from when she was a child growing up in Austria) and against DNA samples of living relatives of Louis, Queen Anna of Romania and her brother Andre de Bourbon-Parme. Tests showed that Naundorff and the Bourbon relatives did not share the same DNA. Despite the fact that the 1950 genetic testing disproved Naundorff and his descendants were of any relation to the Bourbon family, his descendants and their supporters still insist he was the true Lost Dauphin.

The second test was conducted in December 1999 by one of the foremost experts in the case of the Lost Dauphin of France, author Philippe Delorme, who penned *Louis XVII: La Vérité* (Louis XVII: The Truth). Delorme organized DNA tests in a hope to end the debate about Louis XVII once and for all. A segment of

the heart muscle and a piece of the aorta were removed from the preserved heart. Two public notaries watched as two tissue samples were taken from the preserved heart of the Dauphin, now hard as wood, for DNA testing. The notaries also witnessed the transfer of the heart tissues and the opening of the sealed envelopes within the laboratories. The tests were conducted in two different studies, one by Jean-Jacques Cassiman, a professor of genetics at Belgium's Louivian University, and the other conducted by Ernst Brinkman of Muenster University in Germany. The 1999 genetics tests were conducted with every effort to use samples free of contamination. The pieces of the heart would be tested using DNA samples obtained for the Naundorff study. They included tissue samples from two maternal relatives of Louis XVII, his aunts Johanna-Gabriela and Maria-Josepha, and a lock of hair belonging to Marie Antoinette. The heart tissue was also tested against the DNA of Queen Anna of Romania and her brother Andre. Both sets of genetic tests showed that the owner of the heart and the Bourbon family descendants shared the same DNA. There is no way to prove 100 percent that the heart in the crystal urn is that of Louis Charles, but most historians generally accept the DNA evidence as proof that the heart is that of the Lost Dauphin. It doesn't make sense that it would belong to anyone else. The only other plausible explanation is that it belonged to Louis Joseph, the first Dauphin. But the heart does not show any signs of tuberculosis, which killed the prince (Jehaes 2001).

Despite the DNA evidence that the heart is that of the lost king of France, many people still maintain that the Dauphin escaped. One reason, according to Delorme, is that people generally prefer a happy ending. Who wants to think of a small child, barely 10 years old, being poisoned, or beaten, or simply so grossly neglected he died from it. It is much easier to imagine that the Simons were really caring people who spirited the Dauphin out of Temple Prison and that he emerged several years later as a German clockmaker or American missionary. No matter which version you want to believe, the evidence put forth in the Delorme study proves that Louis XVII did not escape his cruel fate. He did indeed die in Temple Prison in 1795. Despite the evidence, a small group of critics claim the heart may have been that of Louis Joseph, the older brother of Louis XVII, though there is little evidence to back their statement. Naundorff's descendants still call themselves Bourbons, refusing to accept the DNA evidence as valid. They have even requested that Naundorff's grave be reopened for further DNA testing.

The Lost Dauphin of France was the victim of the French Revolution and the following Reign of Terror that swept through the country, sending thousands to their deaths. The political pawn of various factions, the Dauphin spent two years locked away from his mother and sister in the Temple Prison, subjected to horrendous treatment by his captors. After his death in 1795, rumors persisted that the boy king had actually escaped Paris and was living abroad, waiting to come home to his throne. For years, imposters from all over Europe claimed to

be Louis XVII, causing his sister Marie Therese bittersweet hope that he was alive and his uncle, Louis XVIII, concern he would lose his throne. Investigations into the young king's death brought little evidence of either his demise or survival. Not until the genetics testing in 1999 was there enough evidence for historians to accept that the heart belonged to Louis XVII. In 2004 the French government gave an official funeral for Louis XVII. The crystal urn containing his heart was displayed in a hearse brimming with lilies—the official symbol of the French Crown. Trumpets blared as the heart of Louis XVII was placed in the royal crypt of Saint Denis Basilica, between the remains of Louis XVI and Marie Antoinette. Finally, the heart of the king, so long one of the most enduring controversies in European history, was laid to rest.

References and Further Reading

Cadbury, Deborah. *The Lost King of France: How DNA Solved the Mystery of the Murdered Son of Louis XVI and Marie Antoinette*. New York: St. Martin's Griffin Press, 2003.

Castelot, Andre. *Queen of France: A Biography of Marie Antoinette*. New York: Harper, 1957.

Cronin, Vincent. *Louis & Antoinette*. New York: Morrow, 1975.

Fersen, Axel Von. *Rescue the Queen: A Diary of the French Revolution*. London: G. Bell, 1971.

Hardman, John. *Louis-Phillipe Memories 1773–1793*. New York: A Helen and Kurt Wolff Book, 1977.

Jehaes, Els, et al. "Mitochondrial DNA Analysis of the Putative Heart of Louis XVII, Son of Louis XVI and Marie-Antoinette." *European Journal of Human Genetics* 9, 3 (March 2001): 185–90.

Lang, Sean. *European History for Dummies*. West Sussex: Wiley, 2006.

Mayer, Dorothy Moulton. *Marie Antoinette: The Tragic Queen*. New York: Coward-McCann, 1968.

Nagel, Susan. *Marie-Therese, Child of Terror*. New York: Bloomsbury, 2008.

13

Charles Darwin got his idea of evolution from "social Darwinist" Herbert Spencer who published first.

PRO   Ian Morley
CON   A. J. Angelo

PRO

The 1800s, a time of substantial industrial growth in Europe and North America, was also an epoch of massive societal transition and cultural upheaval. Industrialization, for instance, affected the political, cultural, economic, environmental, artistic, and philosophical values of many countries, including the principal industrial nations of Britain, France, Germany, and the United States. Moreover, it was a time when certain individuals of undeniable genius and innovative action attained wide levels of recognition that would have been almost unknown to previous generations in the preindustrial context. Engineers such as Joseph Bazalgette and Isambard Kingdom Brunel gained a huge reputation in British society for having helped establish numerous engineering firsts relating to sewerage systems, shipping, tunneling, and bridging. At the same time, architects such as A. W. N. Pugin and artists such as William Morris helped herald new design aesthetics as part of their crusades to reinvigorate the state of British society, in so doing not only influencing many of their vocational associates but also many in society, given the incessant debates about the moral condition of the environment in which people lived and worked on a day-to-day basis. Additionally literary writers, political philosophers, and social commentators like Charles Dickens, the author of *Oliver Twist*; Friedrich Engels, writer of *The Condition of the Working Class in England in 1844*; and W. T. Stead, editor of the *Pall Mall Gazette*, not only engrossed the masses with their musings on contemporary life but moreover to some degree touched on raw social nerves with their realistic descriptions of torpid living conditions for the working classes in places such as London and Manchester, thus helping instigate further discussion about industrial society and its condition. In this way the work of persons like Dickens and Engels was not merely offering sources of amusement to their readers but more so provided both an enlightening and sometimes contentious interpretation of daily life, for they highlighted to the middle classes the harsh realities of day-to-day living in the industrial age of which they, the affluent classes, knew very little. However, arguably the most divisive of all Victorian writers was Charles

287

Darwin, a scientist who promoted a radical thesis of nature's evolution following extensive worldwide travels and research. Darwin from the late 1850s, in particular, created much hullabaloo over his construal of the development of natural organisms and human society.

Traveling from Britain for five years to places such as Argentina, Brazil, Chile, the Galapagos Islands, Australia, New Zealand, Mauritius, and South Africa on the boat the *Beagle*, the Cambridge-educated Darwin accumulated not only a vast amount of knowledge about the world's geology but more so of its plethora of flora and fauna (both alive and dead, that is, from fossils). Darwin published many of his notes on the notion of the characteristics and survival of organisms, and his intellectual apogee was arguably reached in 1859 with the publication of *On the Origin of Species by Natural Selection*, a piece of literature that was perceived as directly challenging the philosophical and theological footing of Victorian society via his outline as to the evolution of nature, and thus humans.

Darwin's attitude and beliefs alarmed many God-fearing members of Western society, drew ridicule from the rather conservative scientific community in Britain and elsewhere, and offended many quarters of the press, which reported rather hysterically that Darwin saw humankind as relations of monkeys. Indeed Darwin's image as created by the mass media in the late 1850s and 1860s led to his satiric depiction as an ape. Darwin's view accordingly drew much disdain, especially from the church establishment, which rejected his line of reasoning instantly, given its belief that nature was created by God as stated in the Old Testament's book of Genesis. Yet regardless of such criticism, Darwin had a monumental philosophical influence by lucidly writing about something rarely commented on and also drawing conclusions that no one had dared print before. So innovative was Darwin's line of thought and so influential was it on scientific and philosophical exploration that the field of biology in subsequent years became dominated by one concept, evolution by *natural selection*, a term introduced by Darwin in *Origin of Species*, although arguably best summarized by the phrase *survival of the fittest*.

Darwin, Spencer, and History

While much controversy was created in Victorian society by the nature of Darwin's thoughts, knowledge, and conclusions, the purpose of this section is not to analyze the disgust of many Victorians toward his ideas but instead to focus on a mystifying matter of historiographical disagreement, one that centers on Darwin's views on evolution not being truly innovative and original. In other words, Darwin's ideas, accepted widely as being original, may be said to not be his own but were actually copied from others. Of particular significance to this view is the manner in which history has been written and understood that has, in effect, meant that many ideas created by Darwin's peers have been subsequently accepted as

belonging to Darwin himself. In this sense, this section reveals how history has rather inaccurately been written with regard to the subject of natural evolution, and that the debate on the theory of evolution should not be seen to belong exclusively to Charles Darwin but should instead place greater reference to other British individuals such as geologist/publisher Robert Chalmers, author of the influential *Vestiges of the Natural History of Creation* (1844), biologist/anthropologist/explorer Alfred Russell Wallace, and, in particular, Herbert Spencer.

Spencer, it should be noted, was a philosopher and political theorist who wrote the powerful piece "Progress: Its Law and Cause" (1857) at the time Darwin was composing *Origin of Species*, and who coined the term "survival of the fittest," a phrase frequently commented on as originating from Darwin. And as such it is arguable that Darwin's conclusions on evolutionary theory leaned heavily on the ideas of Herbert Spencer, in so duplicating many of the points originally set down by his peer. Nonetheless, regardless of this slant, the purpose of this section is to neither denigrate Herbert Spencer's nor Charles Darwin's work. The latter, as previously stated, made a monumental effect on science and social thinking in Britain and European and North American society from the 1850s.

The *social Darwinist* maxim survival of the fittest that is often referred to as belonging to Darwin himself was actually coined by Spencer in reference to the social and cultural changes in the 1800s and the perceived sense of societal advancement borne from the Industrial Revolution. Put into print in 1864 within Spencer's monograph *Principles of Biology*, that is some five years after Darwin's *Origin of Species* (1859), the term, a metaphor for the nature of natural evolution, by its very nature and time of publication evidently leaned toward Darwin's work, which was published some years before, but two points should be noted. First, the idea of endurance by the strongest, as developed by Spencer, was established many years before Darwin put his thoughts on evolutionism into print in the late 1850s and can be found in works such as Spencer's 1857 composition "Progress: Its Law and Cause," and second, Spencer was very precise as to the proper meaning of the phrase. To Spencer the term directly applied to the providence of the wealthy and the poor in the context of a laissez-faire–dominated capitalist urban-based society (Spencer 1857).

In his *Principles of Biology*, Spencer wrote, "This survival of the fittest, which I have here sought to express in mechanical terms, is that which Mr. Darwin has called 'natural selection,' or the preservation of favored races in the struggle for life" (Spencer 1864). However, as previously noted, if the crux of the debate as to the idea of natural selection/survival of the fittest belong to Darwin or Spencer and focuses solely on the date of the concept appearing in published form, then the originator of the allegory on evolutionism quite clearly is Darwin. Then again, Darwin's pioneering work borrowed from Herbert Spencer to such a level that Darwin could be perceived as being a metaphor for Spencer's system of explanation regarding the subject of evolution and organisms in their

environment. As the philosopher Alexander Bain noted in a letter to Spencer in 1863, Spencer was the father of the philosophy of evolutionism, albeit with Darwin supplying the most significant part of the chain. In light of Bain's statement, why has Spencer become so overlooked in societal debates on evolution? To answer this, attention has to turn to Spencer's image as painted by history.

Herbert Spencer's picture as painted by history is, to put it bluntly, far from positive. Not only has Spencer been largely forgotten by mainstream society, and thus from the process in which widely held social history has been written, but when he is referred to, he is frequently painted in such a manner that his image has been smeared as well. To illustrate this point, it is worth considering how the term survival of the fittest, coined by Spencer, has become maligned. To comprehend this, it is imperative to be aware of not only the rise of political correctness in the late 1900s but also of the term's association with the field of eugenics, a line of thought that has connotations with the racial segregation in the 1800s and 1900s and the political strategies and horrifying actions of the nazis in Germany in the late 1930s and early 1940s against Europe's Jews and gypsy population.

On the one hand, the term has been directly connected to a dark time in world history. On the other hand, the positive effects of Spencer's term have been overlooked. By way of example, Spencer and his philosophies were vital in shaping the attitude of one of the world's greatest philanthropists, Andrew Carnegie, described on the Public Broadcasting Service Web site as formerly being the world's richest man, who donated the vast majority of his wealth to support civic causes like education, pension funds, and peace. By the time of his death, Carnegie had donated a sum akin to about $4.3 billion in today's monetary terms. Carnegie also authored the influential work *The Gospel of Wealth* in 1889, a work that softened the perceived harshness of social Darwinism through means of encouraging the rich to involve themselves in benevolence and charitable trusts to assist the daily lives of those less well-off.

The Debate

Social Darwinism is widely used in academic studies and discussions to describe the notion that concepts/theories of biology can be broadened to relate to the social realm and, in so doing, help to explain the development and disposition of society. In its most basic form, social Darwinism repeats a suggested biological principle that competition between organisms propels developmental transition, albeit evidently in the context of society given the sociological appreciation of the concept. This idea though, as commonly shown through the mention of Darwin and Spencer, is not new and can be seen in many respects to predate the 1850s and 1860s work of Darwin and Spencer. In other words, Darwin and Spencer, frequently penned as being the fathers of social Darwinism, commented on a concept that was somewhat historic. By way of illustration, the demographer

and political economist Thomas Malthus (1989) touched on the idea in his extensive publications on population transitions before the 19th century with regard to the struggle for existence.

The influence of Malthus on Darwin should not be underestimated. In fact so strong was Malthus's effect on Darwin that direct reference was made to him in *Origin of Species*, and Darwin himself explained that his work was written in compliance with the doctrines laid down earlier by Malthus. Significantly too, in formulating ideas on social evolution that later became integral to the debate/subject known as social Darwinism, Malthus was not alone. For instance, before the 1850s, that is, the time when Darwin and Spencer were finalizing their theses on nature's development, other writers put forward suggestions about the character of biological and social evolution. One such writer of note was Robert Chambers, author of the highly controversial 1844 publication *Vestiges of the Natural History of Creation*. Like Darwin's subsequent *Origin of Species*, Chambers's work was widely read yet was labeled as containing dangerous ideas, a transgression with which Darwin was tarred after *Origin of Species* was published. Chambers's book implied that God was not the agent actively sustaining the orders of nature and human society. This was perceived as threatening the social order of society and was believed to provide intellectual ammunition that would create mass moral confusion and, as geology pioneer and Fellow of the Royal Society of London for the Improvement of Natural Knowledge Adam Sedgwick predicted, challenge the entire moral and social constitution of the Western world.

Although it is easy to dismiss the influence or shock of debates that make reference to evolutionism, and in so doing social Darwinism, it is nonetheless imperative to view, from the perspective of the early 21st century, the overall discussion on nature and society in an empirically specific manner. In the context of the 19th century, the issue was not so clear-cut. Indeed the debate could

---

Robert Chambers

A Scottish author, Robert Chambers (1802–1871) was especially interested in geology, which he wrote about in several books and devoted attention to in his travel narratives, which focused not on cultural issues so much as geologic explorations and points of interest. In 1844, the year he joined the Geological Society of London (whose previous presidents included Charles Lyell, who had popularized the idea of uniformitarianism), Chambers anonymously published *Vestiges of the Natural History of Creation*. This book was what we would now call a popular science book, drawing on the discoveries in the field and unifying them into a theory that the entire universe had evolved from more basic forms. Along the way, Chambers embraced and forwarded a Lamarckian theory of biological evolution, knowing such views were controversial. Chambers attributed such processes to the work of God, though not as evidence against a theological origin of the universe.

mean a manner of different things to a plethora of people in light of their moral constitution, scientific knowledge, and attitude toward religion, and despite the name of the concept implying that its source of origin was Charles Darwin, it should be reemphasized that the basic idea of social Darwinism, in fact, pre-dates the work of Darwin. Interestingly the term "social Darwinism," despite direct reference to Charles Darwin, was not coined when *Origin of Species* was first published, that is, when debate about evolutionary theory and society was at a crest, but came into being in 1880 in France as *Le Darwinisme Social*. If truth be told, its usage in English was extremely limited in the 19th century and did not filter at all into the domain of popular culture until the mid-1940s, when Richard Hofstadter, an American historian, produced the seminal work *Social Darwinism in American Thought* (1945).

While it is not my intention to break the true meaning of the term social Darwinism down into small pieces, it is necessary to highlight certain key traits that demonstrate the influence of Herbert Spencer on Charles Darwin. This includes, for example, the role of natural selection, that is, a process of nature by which only the organisms best adapted to their environment tend to survive and transmit their genetic characteristics. When applied to the arena of human society, the process transforms itself to explain the advantage particular individuals or social groups have over others as a result of their genetic virtues. This advantage may express itself in monetary terms in that one social grouping may have financial assets and wealth beyond that of their counterparts. Consequently the wider debate may include matters relating to capitalism, which Spencer gave much attention to, plus the rewards and social advantages it bestows on the owners of industry as contrasted with the numerous inconveniences, hardships, and even burdens presented to the laboring population. Seen in light of a natural order, that is, evolution and natural selection, the strongest are able to survive and success is deserved, whereas poverty is a result of being, among other things, "unfit." Such a view can also be employed to explain the status of particular social groups, such as ethnic communities within society, and more precisely can be operated as a tool to explain racism. Hence it—and its axioms such as Spencer's survival of the fit-test—can be criticized for having a subtext attached to race.

Historical Paradox

It is imperative here, in the face of the supposition that Darwin borrowed from Spencer, to investigate the possible influences on Darwin at the time his published work was composed. For instance, it is known that Darwin was aware of Spencer's ideas. In a letter written in 1870 to zoologist and biological theorist E. Ray Lankester, Darwin made clear his esteem for Spencer: "by far the greatest living philosopher in England; perhaps equal to any that have lived" (Darwin 1887: 301). Moreover, in the preface to *Origin of Species*, Darwin's regard for

the theoretical notions developed by Spencer are made clear. In the foreword, described in *Origin of Species* as a "brief sketch of the progress of opinion on the Origin of Species," reference is made to individuals who were for all intents and purposes Darwin's philosophical predecessors. In the preface of *Origin of Species*, Darwin comments on Herbert Spencer's work, composed in 1852, which is work formulated seven years before *Origin of Species* was first published, with the theories proposed by Spencer described as being produced with great energy, and importantly, ability. At the very least, therefore, it seems that Darwin had an acute awareness of Spencer's thinking while he was composing his own understanding of evolutionary theory and was in the build-up phase to publishing his *Origin of Species* (Darwin 1859). Moreover, by 1872, when *Origin of Species* was in its sixth edition, Darwin had clearly embraced the works of Spencer on a level equal to his own. For instance, Darwin commented on how his own term "natural selection" could be replaced by Spencer's "survival of the fittest," which moreover offered more convenience in terms of academic debate (Darwin 1872).

To fully appreciate the substance and true meaning of the above-mentioned statement by Darwin, Darwin's own writing and editing process must be considered. While it may be said that Darwin's 1872 edition was merely a sixth reprint of his original 1859 composition, to make such an supposition would, quite simply, be inaccurate, because it is acknowledged that Darwin rewrote each new edition so that they were different from previous ones. The sixth edition of *Origin of Species* was substantially different from the first version. The importance of this point is central to the argument that Darwin openly imitated intellectual notions composed by Herbert Spencer about the theory of evolution. From 1859 to 1872 it is known that Darwin honed his own interpretation of evolutionism and, as a result, rewrote his ground-breaking work numerous times. With each new edition, he rewrote existing sentences, deleted others, and added new ones.

On this very subject, the historian Morse Peckham (1959), in *The Origin of Species by Charles Darwin: A Variorum Text*, calculated the proportion of change and, on average, with each new edition he estimated Darwin changed 7 percent of his original work. Thus the second edition was 7 percent
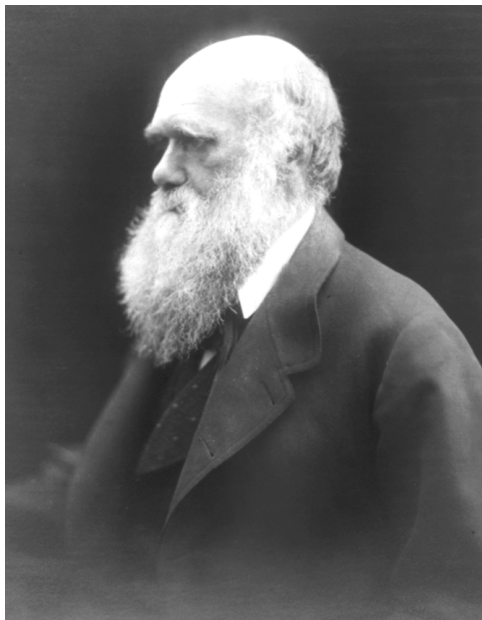
Charles Darwin, shortly before the publication of his controversial book *The Descent of Man* (1871). (Library of Congress)

different from the first edition, the third edition 14 percent different from the first version, and so forth. In light of Darwin's modifications, it may be stated that the sixth version contained many sizable and weighty differences in the mention of ideas and people from the initial 1859 text.

To illustrate this point, text from the 1859 and 1872 editions of the chapter "Recapitulation and Conclusion" can be compared. In the first edition, Darwin wrote:

> In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation. Light will be thrown on the origin of man and his history. (Darwin 1859)

In the second edition, the original text was changed to:

> In the future I see open fields for far more important researches. Psychology will be securely based on the foundation already well laid by Mr. Herbert Spencer, that of the necessary acquirement of each mental power and capacity by gradation. Much light will be thrown on the origin of man and his history. (Darwin 1872)

Evolutionary Individuals

The early phases of Herbert Spencer's and Charles Darwin's careers in many respects were both similar and dissimilar. One major difference was Darwin's five-year surveying experience on board the *Beagle* as it traveled around the world documenting and collecting flora and fauna. Spencer instead was not privy to such wondrous geologic and biological hands-on experience. Nonetheless Spencer gained much practical geologic education while working as a civil engineer for the booming rail industry. While both the young Spencer and Darwin were forging their careers and to some extent augmenting their knowledge of the natural world through very different means, both encountered the same important social and scientific publications. Included in the collection of materials read by both Spencer and Darwin was *Principles of Geology* by Charles Lyell (1830), an influential work that presented a new understanding of rock layering, uniformitarianism as a scientific philosophy, volcanoes, and nature's evolution. Like others about them, Darwin and Spencer were moved by Lyell's composition and, as a result, approached the subject of nature and its evolution in an academic manner for the first time, by this means aiding both men's pursuit to be professional writers and thinkers. Of note, though was, the fact that their approaches to science and matters like nature's and society's evolution were greatly different.

Darwin, given his education and early career work experience, was interested in nature's evolution from the perspective of biology. Spencer's interest

in evolution was rooted in comprehending human social progress, given the intense debates about the Poor Law, a rudimentary form of welfare state, and his schooling by his uncle, Thomas Spencer, a radical reformist clergyman. Important as well was the fact that just as Darwin and Spencer were both swayed by Lyell, they were also both affected by Thomas Malthus's *An Essay on the Principle of Population* (1798). Such was the effect of Malthus on Darwin that in his autobiography, he proclaimed that Malthus provided him with a theory by which to work.

In this respect, Lyell and Malthus, as well as other authors, it should be said, had built a conceptual framework within which persons such as Spencer and Darwin could place their own developing ideas about nature—and about society as well—at a later date. In the case of Darwin, following his extensive vocational travels, experiences, and readings, all that was arguably needed to solidify his emergent abstraction of nature and society was a compelling metaphor that could soundly illustrate the core of his ever-expanding biological and social philosophy. Here then is the value of Spencer in the image-driven vocabulary of Charles Darwin.

The issue of evolution in nature is a complex one, as is the composition of social Darwinism. The formation of social Darwinism, for instance, not only came about as a result of biological erudition but of social, political, and psychological wisdom as well. To this end, knowledge of social science, political theory, the mind, and biology had to be expertly woven together to produce works on the development of society. While both Spencer and Darwin were pivotal in opening up their field, their backgrounds, schooling, and vocational experiences brought them to the same question, that is, evolution, but from entirely different angles. Hence it may be mentioned that Spencer, with his dexterity in comprehending laissez-faire economics, for instance, was more knowledgeable about social science than Darwin was, at least up to the early 1850s, as his 1851 publication *Social Statics* reveals. Spencer devised and employed the term fitness as an adjective to explain the ability of people, and thus social groups, to adapt to unfolding social circumstances, such as the conditions of urban life following the Industrial Revolution and the rise of capitalism. It was this term, replicated continually throughout the 1850s and early 1860s, that led to Spencer's coining of the phrase survival of the fittest in 1864, a term frequently seen as stemming from Charles Darwin, who instead invented the axiom natural selection to describe people's ability, just like flora and fauna in nature, to endure or perish in their natural environment. However, as Darwin found from firsthand experience, his use of the word selection in terms of describing the evolution of nature, people, and society was academically problematic, given its rather ambiguous meaning, which could easily be negated by drawing on Spencer's work on fitness and survival.

Thus it may be said that Darwin not only consumed some of Spencer's ideas but arguably that he also exploited them to make his own theory stronger.

So, did Darwin copy Spencer? There is evidence to suggest Darwin not only had a familiarity with Spencer's work but, moreover, Spencer provided ammunition lacking within the milieu of Darwin's own social analysis. There is ample proof, too, that Darwin respected Spencer's work, and as this section has demonstrated clearly, there are strong associations between the writing of Darwin and that by Spencer. For example, Spencer wrote about the psychological element of human development in his 1870 book *Principles of Psychology*, which explained human evolution in terms of time, space, and causality. Moreover Spencer, an investigator of mathematical descent theory, purposefully sent Darwin a copy of *Principles of Psychology* because he was known to be working on the idea at that time. And Darwin highlighted within the 1872 edition of *Origin of Species* his belief that psychology, in the future, would be founded on Herbert Spencer's work.

In other respects, too, the nature of Darwin's ideas, for example, about understanding human morality, leaned on the work of others, such as Herbert Spencer, even though he and Spencer had fundamentally different approaches to the subject. For instance, Darwin accounted for moral behavior in terms of community selection. Spencer explained human morality in terms of how people inherit and acquire particular social attributes. Even so, when necessary, the nature of Darwin's line of reasoning meant it could draw on Spencer's explanation of human morals. One such case in point was Darwin's rationalization of the large brains humans have in relation to their animal counterparts, a result postulated by Darwin of humans' need to obtain language, thus making the brain evolve into a highly complex organ that would be heritable over time and would grow with the complexity of linguistic communication. Such reasoning, as noted before, at the very least leaned toward people such as Herbert Spencer.

Conclusion

The crux of this work has been to demonstrate how Charles Darwin made use of ideas from Herbert Spencer. Darwin himself, given the nature of evolutionism and his own ideas on the subject, was open in many ways to "using" Spencer's ideas. At the same time, the analysis offered herein has revealed that, in light of Darwin's influence on mid-to-late society of the 1800s, many ideas associated with the idea of natural and social evolution have come to be passed as ideas borne from Darwin when, in fact, they were composed by others. Thus writers and philosophers like Herbert Spencer have become victims of the writing of historiography about evolutionism. But this is not a call for social Darwinism to be replaced by the phrase social Spencerism, because the claim isn't being made here that Darwin copied all of Spencer's work. Instead, first, it is being noted that Darwin leaned on Spencer's work to solve the failings in his own abstraction of evolution and, second, it would be wholly inaccurate to state that Darwin did not know of

Spencer's work nor did not use any of it as well. At the very least, both Darwin and Spencer relied on the same methods and mechanisms to explain matters like human evolution, for example, in terms of mental or moral development, and so some degree of imitation was thus not just possible but in effect likely. At the same time, neither Darwin nor Spencer produced theories of such strength that they were immune from criticism. Both Darwin's and Spencer's ideas showed flaws and relied on the work of other writers and researchers, such as their peers, that is, each other, to iron out weaknesses in their own arguments. Both in their own ways established ground-breaking concepts with regard to explaining the evolution of nature and society, but significant to this process is the writing of history, which has devoted much more attention to Charles Darwin than to, and thus largely at the expense of the reputation of, Herbert Spencer.

References and Further Reading

Athenaeum Collection of Herbert Spencer's Correspondences. A. Bain to H. Spencer. Letter of 17 November 1863. MS 791 no. 67. University of London Library.

Blackledge, Paul, and Graeme Kirkpatrick, eds. *Historical Materialism and Social Evolution*. New York: Palgrave Macmillan, 2002.

Chambers, Robert. *Vestiges of the Natural History of Creation*. London: John Churchill, 1844.

Darwin, Charles R. *On the Origin of Species*. London: John Murray, 1859, 1872.

Darwin, Charles R., with Francis Darwin, ed. *Life and Letters of Charles Darwin*. London: John Murray, 1887.

Hofstadter, R. *Social Darwinism in American Thought*. Boston: Beacon Press, 1945.

Livingston, J. A., and L. Sinclair. *Darwin and the Galapagos*. Toronto: Canadian Broadcasting Corporation Publications, 1967.

Malthus, T. R. *An Essay on the Principle of Population*. Cambridge: Cambridge University Press, 1989.

Peckham, Morse, ed. *The Origin of Species by Charles Darwin: A Variorum Text*. Philadelphia: University of Pennsylvania Press, 1959.

Public Broadcasting Service. "The Richest Man in the World: Andrew Carnegie." Available at www.pbs.org/wgbh/amex/carnegie/ (accessed June 15, 2010).

Spencer, Herbert. *Principles of Biology*. New York: Appleton, 1864.

Spencer, Herbert. "Progress: Its Law and Causes." *Westminster Review* 67 (April 1857): 445–65.

CON

Much has been written about Charles Darwin: his 19th-century life, his work as a naturalist, and his legacy in modern biology. Every year, as regular as the change in seasons, a new crop of Darwin biographies is published. These biographies add to the impressive number of studies already completed on him. They suggest new ways of considering what we already know about his life. They revisit his work in the field and in the laboratory. They survey the influence he's had on modern biological thought. These studies raise and attempt to answer questions central to his life and thought. One question that continues to fuel research interests, scholarly controversy, and speculation is how did he come up with his ideas about evolution?

I will argue that Darwin did not get his ideas about evolution from Herbert Spencer. It is clear why some might argue the opposite. Spencer published two works in the 1850s that suggest similarities between the two thinkers. Spencer, a sociologist and philosopher, came out with "The Development Hypothesis" (1852) and "Progress: Its Law and Cause" (1857) before Darwin published his *Origin of Species* (1859). In the first publication, Spencer discusses the rationality with which to assume that species were created through a process of "modification." This is far more rational, he suggests, than to state that all species appeared as a result of "special creations." The second work highlights the order of organisms: that individual species developed from simple to complex through a process of differentiation. Both works by Spencer explain that such a process as evolution exists and that it's driven to some extent by the forces of competition. On the surface of it, there appears to be an issue of priority: Did Darwin draw his ideas about evolution from Spencer? Spencer, after all, did publish first.

There were many important influences that played a role in Darwin's thinking about biology and transmutation. His family, college experiences, voyage on the *Beagle*, circle of scientific colleagues, and wide reading all contributed to the formation of his ideas. Spencer, however, did not have these experiences. Before turning to direct influences pertinent to Darwin's life and thought, let's first consider the background on evolution.

Background on Evolution

Theories about evolution were around long before Darwin published *Origin of Species*. As discussed in Bertrand Russell's *The History of Western Philosophy*, one of the earliest known versions came from Empedocles, a philosopher of ancient Greece who made his mark before Socrates, Plato, and Aristotle. He suggested that in the first stages of the evolutionary process, there were body parts and organs that coalesced together, as in some primordial soup, sometimes in well-ordered ways and other times not. These "countless tribes of mortal

creatures were scattered abroad endowed with all manner of forms, a wonder to behold" (Russell 2004: 61). Some forms had the bodies of animals and the faces of men. In other cases, the reverse was true. The end result of the evolutionary process, according to Empedocles, has left us only with the creatures that were able to survive.

Ideas about evolution continued to appear just before Darwin's own era. There were plenty of well-developed theories of transmutations that could have influenced his thinking. Parisian scientists at the turn of the 19th century had much to say on the matter. Etienne Geoffroy Saint-Hilaire, student of the great taxonomist Georges Cuvier, allowed for subtle kinds of transmutations. Species, to his mind, had basic forms. These basic forms did not change in their fundamental nature, but at times they did deviate or degenerate into weak, short-lived versions of the true nature. Saint-Hilaire did not believe in the idea that all species descended from a common ancestral origin. A more famous Parisian was botanist and zoologist Jean Baptiste Lamarck. In *Philosophie Zoologique* (1809), Lamarck advocated the idea of evolution and proposed some specific details about how it worked. All animals face pressures from their environment. As environmental pressures change, animals respond with new needs that prompt an increase or decrease in the use of certain organs. Organs or capacities that are increased in use are strengthened and enlarged; those that face decreases are weakened and atrophy. Whatever traits parents have acquired in their lifetimes are passed on to their young. Historians of science refer to this as Lamarck's idea of the inheritance of acquired characteristics.

Even closer to home were the writings of Charles's grandfather, Erasmus Darwin. In *Zoonomia* (1794–1796), Erasmus focused on medicine and other insights he had gained from his medical training and work as a physician. Folded into the pages of this medical text was a concise statement of evolutionary thought. As noted in Janet Browne's *Charles Darwin: Voyaging* (1995), Erasmus

---

Jean-Baptiste Lamarck

A French scientist, Jean-Baptiste Lamarck (1744–1829) was an early evolutionist and an influential naturalist who was one of the first to use the terms biology and biologist to refer to his field. Lamarkian evolution was defined by two laws. The law of use and disuse said that the frequent reliance on any trait would strengthen it, while the disuse of such traits would deteriorate them (accounting for the poor or nonexistant eyesight of underground animals, for instance). The law of acquired characteristics, since disproven but demonstrating a key trend in evolutionary observations, said that new traits acquired by both parents would be passed on to the offspring—Lamarck's essential mechanism of the evolution of a species, as opposed to the simple adaptation that can occur within an individual's lifespan.

spoke of the possibility that "all warm-blooded animals have arisen from one living filament" (E. Darwin 1794–1796). This was essentially the belief of common ancestry. He also added a vague statement about the capacity of animals to change by their own activity or will. This shared something in common with Lamarck's idea about needs.

Shortly before Charles Darwin or Herbert Spencer made public their claims, a work by Robert Chambers called *Vestiges of Natural History of Creation* (1844) laid out yet another version of evolutionary thought. Although *Vestiges* spoke broadly to topics in the sciences—from geology to astronomy, from the origins of life to the origins of language—it received the most attention for its conclusions about evolution. Chambers, who published the work anonymously, understood that the dominant paradigm in biology had long been creationism. The creationist view suggested that a Creator had produced all known species, one by one. What Chambers claimed, from the safety of anonymity, was that a Creator had allowed one species to give rise to another. In other words, evolution was the vehicle of "Providence." While popular in terms of sales, *Vestiges* and other versions of evolution produced few changes in biological thought (Chambers 1844). Biologists of the early 19th century generally held to a form of special creationism when explaining the origin of species.
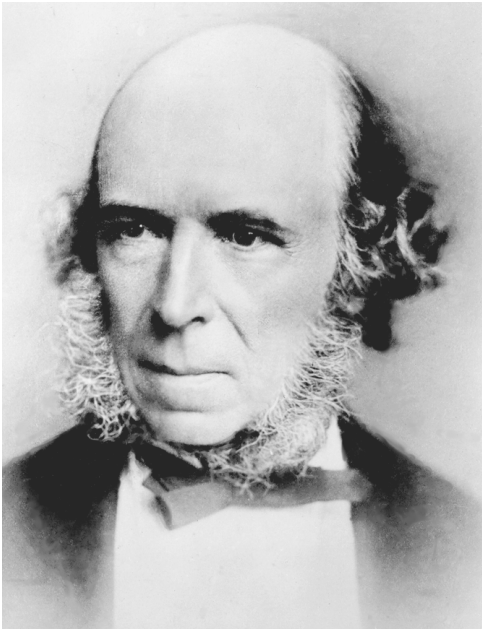
British philosopher Herbert Spencer was the leading proponent of social evolution, later known as social Darwinism. He coined the phrase "survival of the fittest." (Library of Congress)

Neither Darwin nor Spencer, therefore, can be credited with the idea of evolution. Many others had come before them, arguing, advocating, writing, lecturing, and in other ways passing on ideas about how species and organisms might have developed as a result of modifications and transmutations. The distinguishing character of Darwin's work, however, was not in presenting to the world a repetition of what had come before. What differentiated him from previous evolutionary theorists as well as Spencer was twofold: the data-driven nature of his version of evolution and the metaphor he used to explain how it worked. Darwin believed, accurately, that all previous theories of evolution lacked a substantive discussion of the evidence to support the idea. What was needed was evolution grounded in data. His *Origin of Species*, several decades in the

making, aimed to overturn objections by way of evidence. Darwin was also dissatisfied with previous explanations of the mechanism behind evolution. He wanted a law of nature to explain the origin of species the way scientific laws explained other natural phenomenon. If astronomers had successfully explained the movement of the stars and planets without reliance on divine intervention, why couldn't biologists explain the origin of species in a similar fashion? This question, central to understanding Darwin's contribution with the idea of natural selection, arrived later in his life. First, and most critically, were the primary influences and life-shaping experiences.

Family Influences and College Experiences

Darwin's father, Robert Waring, influenced Charles's scientific habits of mind. Robert was a physician who took his son along with him when completing rounds. This gave Charles practice with keeping detailed records of patient histories and case studies. Careful note taking and painstaking record keeping became a hallmark of Darwin's approach to science. In addition, Robert shared science texts with Charles, encouraged the boy's interests in chemistry, and kept a garden filled with rare specimens that stimulated the family's botanical interests. When it came time to leave for college, Robert had left a lasting mark on his son; Charles left to pursue medical training at Edinburgh University.

At Edinburgh, Darwin first tried his hand at medical studies in the mid-1820s. Coursework in anatomy, surgery, chemistry, and related topics gave him a start in the field. But he soon developed a strong distaste for surgery on cadavers and live patients. Lecture demonstrations at the time were conducted without anesthesia. The shrieks of agony and the sometimes careless butchering of individuals proved too much for Darwin. He turned to quiet reading in natural history and resolved to quit medicine altogether.

As he turned away from medicine, Darwin began working closely with Edinburgh professor Robert Grant. Having studied anatomy and embryology in Paris, Grant returned to the university as a lecturer on invertebrate animals. During his training, he had developed an interest in evolutionary theories. Grant shared his excitement about transmutation with Darwin, whom he believed was a kindred spirit of sorts. The primary reason for this was Darwin's family connection to grandfather Erasmus, known to some as the "English Lamarck." Darwin later recalled that he had not yet by then developed any substantive interest in evolutionary thought. Aside from casual reading and awareness of the work of his grandfather and others on the topic, he had no serious inclinations toward the view at that time.

Grant nevertheless took Darwin to meetings of the Wernerian Natural History Society, introduced him to a prominent circle of scientists, and shared a program of research with him. At some point among all these interactions, Grant

revealed his evolutionary stripes to Darwin. This was a grand admission, because evolutionary thought of the mid-1820s was linked to radical views on politics and religion. The British context made this action somewhat equivalent to confessing affiliation with a radical party or admitting to belief in atheism. Darwin, in his *Autobiography*, later recalled this admission and mentioned that he "listened with astonishment," but that the interaction was "without any effect on my mind" (Barlow 1958: 13). At first blush, it's surprising that Darwin gave little credit to Grant for exposing him to evolutionary lines of thought. This was the first scholar he had contact with who made serious claims about transmutation. But historians often turn to a priority dispute between teacher and student as the reason for this diminution in Darwin's recollection. As the story goes, Darwin followed a line of research encouraged by Grant. It had to do with microscopic organisms attached to seaweed. When Darwin made a discovery related to the movement of the organism's ova, he went immediately to Grant with the good news. Grant, in turn, published the findings without mention of Darwin's contribution to the work. Jealousy and sour grapes appear to have brought the relationship to a close. One thing is certain about their time together: Darwin learned a great deal about the process of scientific research, discovery, and publication. The influence, therefore, however murky on the evolution issue, certainly provided Darwin the intellectual tools with which to conduct his own investigations. He had developed the confidence to work independently as a naturalist.

Voyage of the *Beagle* and Circle of Scientific Colleagues

The opportunity of a lifetime presented itself to Darwin at just this moment. He was invited on what is now commonly referred to as the Voyage of the *Beagle*. At the age of 22, he joined the crew of the *Beagle*, a ship that ventured across the Atlantic on a five-year (1831–1836) adventure. The expedition gave Darwin an opportunity to explore the flora and fauna of South America and the Galapagos Islands.

Darwin worked as a naturalist while on the expedition, collecting specimens that would figure prominently in *Origin of Species* over two decades later. He gathered fossils wherever they went. He made notes and drawings in his notebooks about the birds, fish, and other animals he saw. He did the same with flowers, trees, and shrubs. Darwin made a point to use all of his recording skills and scientific knowledge to aid him in keeping a comprehensive log of the journey.

By the time the *Beagle* reached the Galapagos Islands off the coast of Ecuador, he had begun to consider the process of species differentiation. He kept two diaries, one on zoology and another on geology. These record his observations of such creatures as polyps, the way they reproduce, and the manner in which reproduction plays a role in differentiation. The interest he had in this area can be described as a holdover of Grant's influence and his ideas about transmutation. At the same time,

Darwin's geologic observations appear to adopt the kind of thinking set out in Charles Lyell's *Principles of Geology* (1830–1833). Lyell argued that the present is a key to understanding the past. The geologic processes that we see occurring at present are likely to be the ones present in the past. This argument extended geologic time by orders of magnitude not previously taken seriously. Lyell was asking geologists to dramatically rethink the age of Earth. This rethinking of Earth's age facilitated the growth of Darwin's ideas about how new species arise slowly and as a result of modification of earlier ones. Even though Lyell explicitly rejected the ideas of Lamarck and other evolution advocates, Darwin had begun toying with a connection between gradual processes and the rise of new species. The evidence collected while on the *Beagle* seemed to point in that direction.

The turning point for Darwin came less than a year after his return from the voyage. He brought his collections of specimens to the attention of a circle of scientific colleagues. Some agreed to review his fossils; others took his reptiles. John Gould, a respected ornithologist, helped Darwin sort out the bird collection. What they found was remarkable: 13 different species of finches. They had beaks of different shapes and sizes, but they were clearly all finches. What's more, the labeling system Darwin used for the mockingbird specimens suggested that different species of the same bird inhabited separate closely grouped islands. This appeared to confirm for Darwin what he had been toying with on his voyage: the differences in these species most likely came from gradual development from a common ancestor, rather than through a process of special creation, one by one. At this point, in the spring of 1837, Darwin launched his effort to catalog all of the evidence he had collected on the voyage into a system explained by way of evolution, or "transmutation" as he called it then. "The Zoology of Archipelagoes will be well worth examining," he noted at the time, reported in David Quammen's *The Reluctant Mr. Darwin*, "for such facts would undermine the stability of species" (2006). He had begun working on the evidence vital to his major work, *Origin of Species*. Twenty years before Spencer published "Progress: Its Law and Causes," Darwin had become an evolutionist.

He did not announce this to the world, but he kept private record of how his thinking on the subject developed. These records appear in a series of notebooks begun the year he launched his cataloging efforts. In a tip of the hat to his grandfather's evolutionary work, Darwin titled one of these notebooks *Zoonomia*. His notes are filled with reflections on the reptiles, birds, and mammals he collected on his voyage. It's also filled with statements about how "each species changes." These were bold assertions that conflicted with the accepted view of biological thought of the era. What he knew for certain was that species could not be the rigid, immutable forms biologists had long accepted. Rather, he viewed descent from a common ancestor as the more rational, enlightened explanation for the origin of species. The question that remained for him was how exactly the process worked.

Wide Reading

In 1838 Darwin found the answer to his question: natural selection. He came to this idea, in part, as a result of his wide reading. His ideas about evolution, as well as all of his other scientific contributions, were informed by a lifelong habit of reading widely in many different fields. He first learned this habit from his family. Family conversations typical of landed gentry stretched across the arts, literature, music, philosophy, history, government, natural science, and social thought.

Close to the time he began his notebooks about evolution and the early ideas that led to natural selection, Darwin included insights gleaned from his broad reading interests. He made entries about books on birds, autobiographies of explorers, multivolume biographies of literary figures, studies of South America, and debates in philosophy. One of the most famous of these readings occurred around September 1838. That's when he read Thomas Malthus's *Essay on the Principle of Population* (1798).

Some accounts have tended to oversimplify the influence Malthus's work had on Darwin. They suggest that Darwin's idea of natural selection directly followed his reading of *Essay*. This interpretation does not tell the full story. Malthus, a social and political theorist, argued that there are checks and balances on the growth of human populations. If there weren't, they would grow exponentially. Unchecked growth would result in serious overpopulation. One of the most important checks Malthus identified was starvation. The food supply of any community has limits. There is no such thing as an unlimited stock of food. When the supply is plentiful, populations tend to increase; when there is drought or other forms of disruption in the supply, the result is famine, starvation, and ultimately a decrease in the population. Malthus described other checks on population growth, including birth control, war, extreme poverty, and epidemics. But Darwin was most intrigued by the idea of struggle for food.

As Paul Barrett and colleagues note in *Charles Darwin's Notebooks*, in one of his notebooks, Darwin scratched out comments about "the warring of the species as inference from Malthus" (1987). From this and other comments, Darwin alludes to Malthus's influence on his ideas about evolution. Competition for resources such as food create conflicts between species. Some species succeed over others for these resources. The ones that succeed enjoy increases in their population; those that don't tend to starve and become extinct. It became clear to him why so many species of finches appeared with variations in the size and shapes of beaks. Some populations were better suited for survival than others. Those with traits best suited for survival remained. The rest did not. Changes in the food supply played a role in species transmutation. This insight in the fall of 1838 got Darwin moving several years later toward the idea of natural selection.

Darwin's first use of the phrase natural selection appears in what became a draft of ideas for *Origin of Species* written in 1842. During the four-year

period between his insights from Malthus and the draft, Darwin had continued to read widely—this time in the literature on the domestic breeding of animals. With the principles of artificial selection, breeders produce changes in the size, color, and weight of animals. This had long been practiced by those working with cattle, horses, dogs, and other domesticated breeds. The significant conclusion Darwin drew from his reading of Malthus and the breeding literature is that the variety of species found in the wild were a result of "the natural means of selection." He had produced an analogy between the work of breeders and the process of natural species variation. What occurs artificially among domestic breeding practices can be seen, he suggested, in nature as a result of the warring between species for limited resources. The main differences were the lengths of time used to produce each kind of change. The work performed by breeders would occur quickly. In nature, the same effects arise gradually over remarkably long periods of time.

Although Darwin had the data and the mechanism with which to explain evolution, he hesitated. He was in no hurry to produce a final work for many years, although he had discussed the topic with a close circle of scientists. There are at least three main reasons for what scholars refer to as "Darwin's Delay." First, he knew that previous theories of evolution had fallen flat because of a lack of supporting data. This was something he knew he had over other evolutionary thinkers. To do the topic justice, he believed the final work must be a multivolume encyclopedic magnum opus that would fell all opposition by the sheer weight of evidence. The desire to produce such a work made him wait before publicly announcing his conclusions. Second, Darwin struggled with a lifetime of illness and family losses. He was prone to vomit if he became too excited or agitated. Intense debate or controversy did not agree with his constitution. What he was proposing in his work would upset the norms of standard biological thought. His weak constitution conflicted with his desire to offer a new scientific paradigm. At crucial points during his work, Darwin also suffered personal setbacks. In the late 1840s and early 1850s, his father died followed by the death of his daughter, Annie. Both losses took a toll on his emotional and professional energy.

Third, the publication of *Vestiges* figured into his delay before coming out with his own version of evolution. The reviews of Chambers's work were not favorable, despite the number of copies it sold. The last thing Darwin wanted to do was come out with his theory of evolution at a time when the notion was being dismissed by a critical audience. Darwin knew he had an even more challenging task before him in making the case for evolution without reference to divine intervention, a concession *Vestiges* had made. What Darwin wanted to establish was a law of nature that governed the origin and differentiation of species. Back he went to the idea of a magnum opus of endless proportions.

He might have gone his entire life without publishing *Origin of Species* if not for a discovery by Alfred Russell Wallace. Darwin had left a clean copy of his draft of 1842 in a file with instructions to his wife on how to go about

publishing it if he were to die an untimely death. He might have relied on this as a precaution in case he never published on evolution during his lifetime. This became unnecessary when Wallace entered the picture in 1858. Wallace was a young naturalist conducting field work in Asia. He collected specimens much the same way Darwin had decades earlier. They shared similar observations, had similar habits of reading widely across fields of interest, and had developed a very similar line of thinking about the origin of species. They both had independently developed a theory of evolution by natural selection.

In an unusual twist of fate, Wallace decided to reveal his discovery to Darwin. In a package delivered in 1858 containing an explanation of the theory, Wallace requested his thoughts and commentary on the matter. Darwin quickly met with his circle of colleagues, including botanist Joseph Hooker and geologist Charles Lyell. The decision was made to sponsor a joint presentation of papers from Wallace and Darwin at a future meeting of London's Linnean Society. While this resolved an issue of priority for Darwin, it left him in a race against time to produce the work he had long intended to write on the subject. When it finally appeared the following year, it did so with the title *On the Origins of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. The work was an intellectual bombshell that is considered a pillar of modern biology.

Conclusion

It would be fair to say that there were many influences pertinent to the development of Charles Darwin's theory of evolution. His father's interests in medicine and botany gave Darwin his first introduction into the wonders of the natural world. Robert Grant furthered his development as a scientist, trained him in actual research, and shared with Darwin a staunchly held belief in evolutionary thought. The voyage of the *Beagle* dropped Darwin in a sea of evidence that hinted at descent from common ancestral origin. Cataloging the specimens he had collected on the voyage gave Darwin the confidence to start notebooks that provide the basis for *Origin of Species*. His wide reading, moreover, helped generate the framework with which placed the evidence for evolution. The result was a theory of evolution unlike any that had come before.

This section has argued that Charles Darwin became an evolutionist in the mid-1830s independent of Herbert Spencer's work. Darwin's notebooks affirm this and, later, his brief draft of ideas written in 1842 became a first attempt at *Origin of Species*. Thus, Spencer may have published two works in the 1850s before Darwin went public with his ideas at the end of the decade. But as most Darwin scholars recognize, the ideas for *Origin of Species* were very long in the making, informed by experiences on the *Beagle* and unique in the book's use of data and the metaphor of natural selection. Spencer, in an acknowledgment to

Darwin's influence on his work, changed his use of key phrases after *Origin of Species* made its debut. In the 1852 publication "The Development Hypothesis," Spencer mentioned the "Theory of Lamarck." When he republished the article after *Origin of Species*, he changed the phrasing to the "Theory of Evolution." As such, Darwin influenced Spencer's writing, thinking, and selection of phrasing. The question, therefore, is not whether Darwin got his ideas from Spencer. Rather, our attention should be directed to the extent to which Darwin influenced Spencer's ideas about society.

References and Further Reading

Barlow, Nora, ed. *The Autobiography of Charles Darwin, 1809–1882*. London: Collins, 1958.

Barrett, Paul H., Peter J. Gautrey, Dandra Herbert, David Kohn, and Sydney Smith, eds. *Charles Darwin's Notebooks, 1936–1844*. Ithaca, NY: Cornell University Press, 1987.

Browne, Janet. *Charles Darwin: Voyaging*. Princeton, NJ: Princeton University Press, 1995.

Chambers, Robert. *Vestiges of the Natural History of Creation*. London: Murray, 1844.

Darwin, Charles R. *On the Origin of Species*. London: Murray, 1859, 1872.

Darwin, Charles R., with Francis Darwin, ed. *Life and Letters of Charles Darwin*. London: Murray, 1887.

Darwin, Erasmus. *Zoonomia; Or the Laws of Organic Life*, 2 vols. London: 1794–1796.

Darwin, Francis., ed. *The Foundations of the "Origin of Species": Two Essays Written in 1842 and 1844 by Charles Darwin*. Cambridge: Cambridge University Press, 1909.

Lamarck, Jean-Baptiste. *Philosophie Zoologique*. Paris: Dentu et L'Auteur, 1809.

Lyell, Charles. *Principles of Geology: Being an Attempt to Explain the Former Changes of the Earth's Surface by Reference to Causes Now in Operation*. London: Murray, 1830–1833.

Malthus, T. R. *An Essay on the Principle of Population*. Cambridge: Cambridge University Press, 1989. Original edition 1798.

Quammen, David. *The Reluctant Mr. Darwin: An Intimate Portrait of Charles Darwin and the Making of His Theory of Evolution*. New York: Norton, 2006.

Russell, Bertrand. *The History of Western Philosophy*. New York: Routledge, 2004.

Spencer, Herbert. "The Development Hypothesis." *The Leader* (March 20, 1852).

Spencer, Herbert. "Progress: Its Law and Causes." *Westminster Review* 67 (April 1857): 445–65.

14

Slavery was unprofitable for slave owners.

PRO   Radica Mahase
CON   Jerry C. Drake

PRO

The profitability of slavery has been exaggerated. There is little doubt that the Atlantic slave trade was profitable to some extent. Undeniably, no commercial venture could last four centuries without making some amount of profit. What is questionable, however, is the amount of profit made. Also questionable is whether the profits of slavery benefited the countries that practiced it or just benefited specific groups such as the slavers and plantation owners. Perhaps the Caribbean historian Eric Williams (1944), who said that slavery financed Britain's Industrial Revolution, has made the boldest claim that slavery was profitable. Other historians, such as Hugh Thomas (1997), hold that slavery's profits were too small to have significantly affected the pace of Britain's economic development. Moreover, the eighteenth century theorist, Adam Smith, who founded the discipline of economics, held that slavery as an institution was inherently unprofitable for the economy as a whole.

The debate has continued for almost a century and a half after the abolition of the slave trade. For the past 50 years, newer historians have brought new sources, but not entirely original perspectives, to the debate. The discussion has centered on both theoretical and empirical issues pertaining to the profitability of African slavery. On one level, it has attempted to determine the extent to which certain countries, societies, groups, and individuals benefited from the enslavement of Africans. On another level, there is the comparative aspect of the debate: an endeavor to determine whether slave labor was as profitable as free labor.

The most influential addition to the historiography has evaluated the extent to which the slaveholders benefited from their investments in slaves and has assessed the viability of slave economies and analyzed the effects of slavery on the growth and development of certain countries and regions. The topic has stimulated many discussions, but the answers provided have been influenced by the manner in which the question is asked and the perspective(s) of the individuals answering the questions. The way in which profit has been defined and assessed is one of the most problematic aspects of writings on this topic. The total profit

309

calculated depends on the economic historian's interpretation of what should or should not be included in the calculations.

The unprofitability of slavery is normally explained by examining the slave trade and the institution of slavery either as an industry or as a business enterprise. Much of the discussion on unprofitability specifically has centered on the extent to which individuals and companies did not benefit from African slavery. The question arises as to whether slavery was unprofitable for the investors, slave traders, and slave/plantation owners. Some studies, including "The Fishers of Men: The Profits of the Slave Trade" by R. P. Thomas and Richard N. Bean (1974), have examined the unprofitability of slavery to African intermediaries and the slaves themselves. Were there any major differentiations in unprofitability and losses experienced for each of these groups? Was the level of unprofitability more or less constant for all groups involved? Who experienced higher economic losses? The existing historiography has attempted to provide answers for these kinds of questions. The main arguments so far have centered on the extent to which slave traders benefited from the slave trade, the level of profits realized by plantation owners, the extent to which African slavery stimulated industrial developments in Britain, and the manner in which African intermediaries benefited from the institutional enslavement of the African population.

The main contributors of the debate on unprofitability so far are those who can be referred to as an "older generation" of historians. Roger Anstey (1975b) gave an estimate of the profitability of the British slave trade for over 50 years in his "Volume and Profitability" and calculated that the rate of return from the slave trade was lower than 30 percent, the figure claimed by some other historians. Joseph E. Inikori (1992) argued that the slave trade was not competitive; it, rather, was monopolistically organized and it yielded high profits. Thomas and Bean, conversely, claimed that the trade was a perfectly competitive industry. B. L. Anderson and David Richardson (1983), in "Market Structure and Profits of the British African Trade in the Late Eighteenth Century," also argued that the slave trade was competitive and that it resulted in normal profits, therefore

---

History of the Slave Trade

Slavery had been practiced in Britain since antiquity, but most forms of slavery were gradually outlawed by the Middle Ages, with rare exceptions (enforced servitude was at one point used as a punishment for Irish Catholics, but this wasn't the same as the wholesale purchase and sale of people). When the enslavement of Africans began in the 17th century, it was unchallenged in courts and expressly allowed in a 1729 declaration by the attorney general. Other colonial powers such as the Dutch and the French practiced African slavery at the same time, though the practice was (though vast) short-lived: France outlawed it early in 1794, in the aftermath of the French Revolution, and Britain abolished slavery 40 years later.

profitability was low. Other historians have provided variations of these debates, while some individuals have used new data along with older methods to calculate the degree of unprofitability that existed at various levels.

The Slave Trade

Economic historians have attempted to calculate the profitability of the slave trade in three ways. The first looks at the competition in the market of slaves. The second examines financial accounts of individual slaving ventures. The third attempts to estimate overall profits from total costs and revenues. Economic historian Eric Williams, in *Capitalism and Slavery*, claimed that the slave trade was not only highly profitable but that it also provided the capital needed to finance industrial development in Britain. Williams referred to a profit of 30 percent and stated that the profits from the triangular trade influenced Britain's entire productive system. However, the slave trade was not as profitable as Williams calculated, because he omitted various costs that slave traders had to cover that would decrease profit.

Stanley Engerman (1972), in "The Slave Trade and British Capital Formation in the Eighteenth Century," for example, pointed out that Williams's calculations of high profit is a result of a comparison between the costs of producing the goods traded for the slaves with the prices received for the slaves. However, this excludes costs of shipping, mortality, and other costs. The calculations of Roger Anstey, in contrast, are probably more realistic. He calculated that the rate of return for the British slave trade was 9.6 percent between 1761 and 1807. He noted that the merchant who invested in the slave trade had to wait for approximately one year before receiving any return on his investment.

The slave trade was, in fact, variable, and the profits depended on numerous factors. These included the general risks of the shipping industry, loss of merchants' vessels and cargoes to privateers or by foundering, and unpredictable factors that influenced trading activities. The level of profitability was also determined by the timing of the slaving expedition, market conditions, the duration of the middle passage, and morbidity and mortality. Added to this, there was a high cost of entrance into the slaving business.

Slave shipping was expensive activity. The time needed to outfit a slave ship and the journey itself lasted more than one year. Slavers often had to recruit and pay for a crew double the size of that as other commercial activities to stem revolts on board the ships. Often the insurance was high, because traders needed to cover the value of the ship, the merchandise, and the human cargo as well. They also had to purchase large quantities of food and provisions for the crew and human cargo. Added to this, slave traders were required to purchase various permits, pay many duties, pay fees at each port of call, and pay broker and pilot fees and the wages of the carriers and the boatmen who transported the enslaved Africans and provisions. Some slave traders bought customary presents for

important African rulers or merchants. Most of them had to pay fees for auction-eers, announcers, and duties to officials and churches. There were rumors that various companies were purchasing slaves for about £3 per head and reselling them at £20 in the United States. However, as Hugh Thomas points out in *The Slave Trade*, the promulgators of this rumor did not consider that the companies had to offset various costs such as the management of forts in Africa and the transportation of cargo from Britain.

Furthermore, profit making was not always a constant factor. Richard Bean's (1971) calculation in "The British Trans-Atlantic Slave Trade, 1650–1775" of the estimated profit per slave shows this quite clearly. Bean estimated that for the period from 1651 until 1675, the profit per slave was £14.66. From 1701 to 1720, the profit had decreased to £8.46. The next two decades saw a further decline, and the profit per slave was £3.16.

On a more microscopic level, it can be noted that while some slave ships made heavy profits from a journey, others struggled to break even and some experienced heavy losses. This differentiation in profit can be seen when one ship, which had undertaken three journeys, is examined. According to the Davenport Papers, the British slave ship *Hawke* made a profit of over £7,000 on its first trip in 1779. In 1781 it made almost £10,000. However, on its third journey in 1781, it fell to French privateers. This represented a gross loss higher than the profits made. Other examples of British slave ships experiencing heavy losses can be seen through the discussion of Francis E. Hyde, Bradbury B. Parkinson, and Sheila Marriner (1953) in "The Nature and Profitability of the Liverpool Slave

Deck plan of a slave ship, showing how captured Africans were arranged as cargo, 1700s. (North Wind Picture Archives)

Trade," of the case of the *May* in 1772, which lost more than £1,500, and the *Hector*, which experienced a loss of £3,000 in 1773. Johannes Postma (1990) has calculated that of the Dutch slaving ships, 113 made a minimum profit of 5 percent, 14 broke even, while 32 incurred losses. Of those that incurred losses, one ship experienced a loss of 37 percent and another had a loss of 48 percent. Thomas noted that during the second half of the 18th century, of the 100 Dutch slave ships, 41 operated at a loss, and profits were barely more than 3 percent. According to Thomas, John Newton, once a slaveholder and later an abolitionist, noted that the slave trade could be seen as a sort of lottery in which every adventurer hoped to gain a prize.

Another differentiation can be seen in the type of the slave-trading venture. Profits were made by independent traders more than by national privileged companies. The companies often spent heavy sums on the upkeep of forts, bureaucratic activities, and bribes that were necessary to placate local police and officials at the landing ports. In addition, the staffs of companies often tried to make private profits, which decreased the overall profit of the companies. For example, according to Thomas, the Middleburgh Company made an annual profit of 1.43 percent from 1761 until 1800 as opposed to the 8 percent profit it made from the period 1751 to 1760. In fact, from 1770 until 1800, the Middleburgh Company did not realize any average profit, and its average loss increased progressively. The profits of French slave traders in Bordeaux, Nantes, and La Rochelle, according to Seymour Drescher (1977) in *Econocide: British Slavery in the Era of Abolition*, was not more than 6 percent and often was as low as 1 percent. The larger slave shipping merchants were only able to realize substantial profits because they invested in other financial ventures such as sugar and coffee plantations.

An additional factor determining the levels of profit made was that of the price of a slave. There were distinct variations in the price of enslaved Africans from the 16th to the 19th centuries. According to Thomas, the price of a slave was £5 in the 1670s, but 100 years later an enslaved African was sold for £30 to £50. By the mid-19th century, the price had dropped to an average of £15. Thomas suggested that the decrease in the price of slaves led to considerable damage to the slave trade in general. He also noted that slave traders were rarely paid fully for their cargoes upon delivery, and payments were frequently made in commodities. Thus many buyers purchased slaves on credit, and in the process, they incurred high debts. Profits therefore decreased after the second half of the 18th century as the prices of slaves increased in Africa. In fact, while profits from the slave trade averaged 8 to 10 percent, this rate of profit was the same as for other commercial ventures. Even after 1807 when the slave trade was illegal and the price of slaves increased, the large slaving companies were only able to realize substantial profits when they invested in sugar or coffee plantations. Otherwise they faced the risk of going bankrupt.

The Plantation Economy

According to Keith Aufhauser (1974) in "Profitability of Slavery in the British Caribbean," certain variables are needed before the historian can calculate the profitability of the slaves and slaveholdings within the plantation system. Data pertaining to the productivity per slave, longevity, reproduction and mortality, capital costs and expenditure for the maintenance of the slaves, as well as the process of sugar and rum are necessary for the individual to tabulate the actual profits derived per slave. The limitations in the availability of these types of data have made it difficult to calculate the actual profitability of the slaves within the plantation economy. Nevertheless, historians have attempted to determine the exact extent of the profits realized by plantations. Ulrich B. Phillips (1918), in *American Negro Slavery*, looked at the southern region of the United States. Here, slaveholders did not realize a profit by the end of the first half of the 19th century. The majority of the slaveholders were barely making enough to cover maintenance and other costs, and only those situated in the most advantageous parts of the region were able to make profits. In *Slavery in Mississippi*, Charles S. Sydnor (1965) used statistics for the Mississippi area of the United States to show that there was a low rate of return on a large number of plantations.

The difference between slavery and the plantation economy has been generally ignored from the existing historiography. Slavery has been used interchangeably with the plantation economy, and the plantation economy has been defined, structured, and assessed around slavery. However, slavery was one part of the plantation economy, and the high state of profitability of the plantation economy should not be translated into the profitability of slavery as a business enterprise. Even if slaveholders and plantation owners realized a profit, this does not mean that slavery was economically profitable. Plantation owners were normally engaged in economic activities either totally independent of slavery or slave labor, or engaged in economic ventures that partially revolved around slave labor such as cocoa, coffee, tobacco, and tea.

As L. J. Ragatz (1963) discusses in *The Fall of the Planter Class in the British Caribbean: 1776–1833*, plantation profits often oscillated because of increases in maintenance costs and fluctuations in the wholesale price of sugar. For the plantation to operate at a state of high profitability, the capital invested per slave had to be low. In *The Nature and Properties of the Sugar Cane*, G. R. Porter (1830) estimated that the average investment in buildings and utensils in Barbados and Tortola was £25 per slave in the 1820s. Aufhauser calculated that the capital cost per slave was £176.7. This included buildings and utensils, cattle and horses, land and the slave. Solon Robinson noted that the profits of the plantation of Colonel L. M. Williams of South Carolina were very low. The plantation was valued at $160,402. The expenses amounted to $17,894.48. This included the interest of 7 percent on the investment in land, slaves and livestock,

taxes on slaves and land, medical care, the wages of the three overseers, and the expenditure on tools and equipment. When the income from other sources was added and the average cost of a pound of cotton was derived, Robinson concluded that the plantation was making a profit of about 3 percent. (Govan 1942: 514) Sydnor looked at figures for the profit and loss of plantations in the southern region of the United States and tabulated that the profits of the plantation owners were less than $880 when all expenses had been deducted. Sydnor then concluded that there was a very small rate of return for plantation owners.

Added to this, some slaves were more productive and therefore more profitable than others were. A more blatant example can be seen with female house slaves employed in the great house to cater to the needs of the masters and were therefore engaged in activities more of a sexual nature. These were often less productive than the field slaves who were directly engaged in agricultural production. Another differentiation can be seen within the field slaves, when the productivity of the "great gang" is examined against that of the second or third gang. In addition, differences can be seen with the low numbers of slaves employed as skilled slaves or craftsmen. Levels of productivity and maintenance went hand in hand, and it cost more to maintain slaves who were more productive. An exception to this rule could have been the female slave who was employed in the master's house to provide sexual services to the planter and who was treated in a highly favorable manner. Generally the high cost of maintenance was only met satisfactorily when plantation owners invested in other more lucrative economic activities.

Williams has argued that the decline in the profitability of the plantation economy was a key factor influencing the emancipation of enslaved Africans in 1833. He noted that there was an overall decline in the British West Indian sugar economy by the end of the 18th century and that this, more than humanitarian considerations, led to the abolition of slavery. The emergence of the 19th century heralded an era of unprofitability, inefficiency, and fluctuations in the plantation economy. David B. Davis (1987), in "Reflections on Abolitionism and Ideological Hegemony," noted that African slavery was associated with a state of indebtedness, soil erosion, and deserted properties. In addition, according to David Ryden (1963) in "Planters, Slaves and Decline," planters adopted new management techniques to cope with the increasing cost of plantation supplies and the low prices of commodities on the international market. First, the sugar planters were highly indebted to British merchants and financiers. Second, soil exhaustion and limited or no technological improvements and changes placed West Indian planters in a situation where they were unable to compete with other sugar-producing colonies. Third, West Indian planters were experiencing unfavorable economic competition, which resulted in desertion of some plantations. These ideas are supported by Ragatz, who noted that there were important economic changes occurring in the West Indian plantation economy and that there was a close

relationship between the decline in productivity and profitability and the abolition of slavery.

Capitalism and Industrialization

One of the most advocated "benefits" of the enslavement of Africans is the notion that slavery led to the emergence of capitalist enterprises and economies and the rise of industrialization. Williams has postulated that the profits of the slave trade and the plantation economy were sufficiently lucrative to finance the industrial developments that occurred in Britain from the 18th century onward. He argued that the profits made by slave owners were invested in industries in Britain. However, the Williams thesis (as Williams's arguments became known) omitted the fact that overseas demand for British goods increased because of independent developments outside of Britain. Also, industrial developments in Britain were partially an effect of British demands for imports. According to Drescher, the expansion of purchasing power overseas was transmitted back to British exporters via a network of colonial trading connections. In fact, the actual contribution of profits to British national income was small, usually lower than one-half of 1 percent of the British national income.

Drescher calculated that of the total slave profits, only £14,000 per annum was invested in industrial development in Britain. Thus the slave trade contributed 1.59 percent of the total national income of Britain. According to Anstey, if all profits from the slave trade were invested in industries, it would only amount to 7.94 percent.

It is correct to say that some profits of the slave trade contributed to the development of specific industries as seen in the relationship between the profits of the Liverpool trade and its contribution to the Lancashire cotton industry. However, it is incorrect to say that the slave trade or slavery was profitable to the extent that it funded the Industrial Revolution in Britain.

Also questionable in the Williams thesis is the question of the actual share of the profits invested in British industries, because surely some part of the profits was put aside for consumption purposes, was invested in agriculture, or was reinvested in the slave trade or into the plantation system. Many slave traders and trade holders reinvested some portion of their already meager profits in land and other trading activities. Thus it is hardly likely that the entire profits of the slave trade and slavery were pumped into industries in Britain, and even if this were so, it was still not sufficient to stimulate the level of industrial activities that occurred in Britain. Inikori is probably more correct when he notes that there was indeed a close relationship between slavery and British industrialization. However, this relationship was not as directly related as Williams has made it out to be. A high capital investment from slavery was not directly invested in industries. The relationship was a more indirect one where

the slave trade contributed to the general growth of trade, transport, and manufacturing. As Barbara L. Solow and Stanley L. Engerman (1987) note in *British Capitalism and Caribbean Slavery*, West Indian slavery stimulated British overseas trade, provided new markets for exports, and widened the markets for goods produced by enslaved Africans.

In fact the profitability of slavery as an economic system was often exaggerated, and there were situations where it actually retarded economic growth rather than fostered development in all areas where it existed. A good illustration of this can be seen with the economic backwardness of the southern region of the United States. In this region, slavery proved to be a deterrent to economic growth. The northern region was able to progress economically at a much faster pace than the southern region. In fact the South was characterized by diminishing commerce and retarded economic growth.

Slavery in the American South has been blamed for a lack of economic growth, because slave labor was economically expensive and ineffective. First, slavery hindered industrial and commercial development of the southern region, because capital was frozen in the form of labor and was therefore not available for manufacturing and commerce. Capital was normally tied up in the ownership of slaves and the plantation system so that capital that could have been used to construct factories was concentrated in the ownership of slaves and was inflexible. In addition, planters did not save enough to reinvest in their plantations or in other economic activities. According to Robert Russel (1938), planters lived and operated on the anticipated income from the next crop and not from the income from the preceding crop. Many planters lived on advances, which they obtained from British companies or firms in the northern region. They often used their crops as security so that once there were fluctuations in the prices of crops, they became heavily indebted. As a result, savings were low and capital was not available for investment. As noted by Douglas F. Dowd (1958) in "The Economics of Slavery in the Ante Bellum South," slavery also hindered the diversification of the southern economy so that the South became dependent on the North. Slave owners preferred to engage their slaves in agricultural activities as opposed to manufacturing activities. The use of slave labor hindered the development of a home market for local industries in the South, where local markets were lacking because the slave labor force either paid little or nothing at all or paid in kind rather than in cash. Thus the propensity for the development of a capitalist system was curbed and there were serious implications on consumption-savings relations.

Slavery prevented maximum use of skills and abilities of the labor force. The majority of the enslaved labor force was employed in agriculture rather than as skilled personnel. A social system developed where the white labor force perceived labor as a socially degrading activity and refused to work in this area. Also, the idea that whites should not perform menial jobs such as personal services like cooking and washing became widely accepted and practiced in the

South. There were also situations where whites refused to work alongside African slaves. Consequently the entire labor force in the southern region was underused. Added to this, the use of slave labor may have retarded the implementation of agricultural machinery. This occurred, according to Russel, because slave labor was relatively cheap and labor intensive and was used instead of machinery, and because plantation owners often did not trust their slaves to use expensive, often complicated machines. It became a situation where southern plantations with the most fertile land and efficient supervision were able to profit from the plantation system. In general, however, according to Harold D. Woodman (1963) in "The Profitability of Slavery," slaveholdings in the South were not profitable, although some individual planters made large profits.

Conclusion

The profitability of African slavery has often been exaggerated and overemphasized in the historical literature. When attempting to determine the level of profitability that existed, it is imperative to examine the various factors that influenced profitability as well as the differentiations that existed. It is rather difficult to determine the actual extent of (un)profitability that existed as a result of limitations, sources, evidence, and so on. Perhaps this is why the debate has continued in a somewhat unaltered form for all these years. Some of the statistics used in the available studies are questionable if one considers that officials have collected them. Aside from the bias of these officials, they were not always diligent in their statistical recordings and compilations. Furthermore, for actual levels of (un)profitability to be calculated, specific trends in slave ships and plantations that used slave labor must be examined. However, this has been a rather difficult task for economic historians because of the unavailability of records. It is also necessary to define levels at which (un)profitability have been assessed and calculated, because terms like profits, productivity, and costs have varied meanings to different individuals. It has often been the accepted view that the enslavement of Africans was a profitable enterprise whereby various individuals, companies, and economies realized substantial benefits—but this was not the case.

References and Further Reading

Anderson, B. L., and David Richardson "Market Structure and Profits of the British African Trade in the Late Eighteenth Century: A Comment" *The Journal of Economic History*, pp. 713–21 (v.43, 1983).

Anstey, Roger. *The Atlantic Slave Trade and British Abolition 1760–1810*. Atlantic Highlands, NJ: Humanities Press, 1975a.

Anstey, Roger. "The Volume and Profitability of the British Slave Trade, 1761–1807." In *Race and Slavery in the Western Hemisphere: Quantitative*

Studies. Edited by S. L. Engerman and E. D. Genovese. Princeton, N.J.: Princeton University Press, 1975b.

Aufhauser, R. Keith. "Profitability of Slavery in the British Caribbean." *Journal of Interdisciplinary History* 5, 1 (1974): 45–67.

Bean, Richard Nelson. "The British Trans-Atlantic Slave Trade, 1650–1775." PhD thesis, University of Washington, 1971.

Davis, David B. "Reflections on Abolitionism and Ideological Hegemony." *American Historical Review* 92, 4 (1987).

Dowd, Douglas F. "The Economics of Slavery in the Ante Bellum South: A Comment." *Journal of Political Economy* 66, 5 (1958): 440–42.

Drescher, Seymour. *Econocide: British Slavery in the Era of Abolition*. Pittsburgh: University of Pittsburgh Press, 1977.

Engerman, Stanley L. "The Slave Trade and British Capital Formation in the Eighteenth Century: A Comment on the Williams Thesis." *Business History Review* 85, 4 (1972).

Fogel, Robert William. "Three Phases of Cliometric Research on Slavery and Its Aftermath." *American Economic Review* 65, 2 (1975): 37–46.

Govan, Thomas P. "Was Plantation Slavery Profitable." *Journal of Southern History* 8, 4 (1942): 513–35.

Hyde, Francis E., Bradbury B. Parkinson, and Sheila Marriner. "The Nature and Profitability of the Liverpool Slave Trade." *The Economic History Review* 5, 3 (1953): 368–77.

Inikori, Joseph E., and Stanley L. Engerman, eds. *The Atlantic Slave Trade: Effects on Economies, Societies, and Peoples in Africa, the Americas and Europe*. Durham, NC: Duke University Press, 1992.

Phillips, Ulrich B. *American Negro Slavery*. New York: Appleton, 1918.

Porter, G. R. *The Nature and Properties of the Sugar Cane*. London: Smith, Elder & Co., 1830.

Postma, Johannes. *The Dutch in the Atlantic Slave Trade, 1600–1815*. New York: Cambridge University Press, 1990.

Ragatz, L. J. *The Fall of the Planter Class in the British Caribbean: 1776–1833*. New York: Octagon Books, 1963.

Russel, Robert R. "The General Effects of Slavery upon Southern Economic Progress." *Journal of Southern History* 4, 1 (1938): 34–54.

Ryden, David. "Planters, Slaves and Decline." In *Capitalism and Slavery—Fifty Years Later: Eric Eustace Williams—A Reassessment of the Man and His Work*. Edited by Heather Cateau and Selwyn Carrington. New York: Lang, 2000.

Solow, Barbara L., and Stanley L. Engerman. *British Capitalism and Caribbean Slavery*. Cambridge: Cambridge University Press, 1987.

Sydnor, Charles D. *Slavery in Mississippi*. Gloucester, MA: Smith, 1965.

Thomas, Hugh. *The Slave Trade*. New York: Simon and Schuster Paperbacks, 1997.

Thomas, R. P., and Richard N. Bean. "The Fishers of Men: The Profits of the Slave Trade." *Journal of Southern History* 34 (1974): 885–914.

Williams, Eric. *Capitalism and Slavery*. Chapel Hill: University of North Carolina Press, 1944.

Woodman, Harold D. "The Profitability of Slavery." *Journal of Southern History* 29 (1963): 303–25.

CON

The question of the profitability of antebellum slavery in the United States has remained a source of discordant debate for nearly two centuries. Prior to the abolition of the South's so-called peculiar institution, the argument over the profitability of the slave system was waged in the broader social arena between pro-slavery advocates and abolitionists. Following the Civil War, this question became the subject of much rancor among a variety of scholars, chiefly historians and economists.

At the heart of the debate resides an apparently simple question: Was slavery a profitable institution? What initially appears to be a rather cut-and-dried query is actually far more complicated. What do we mean by "profitable"? Who could be said to have profited from this system? The slave? The owner? The planter class? The South as a region? Or the American nation as a whole?

The profitability of slavery is a difficult question to ponder, for the key to the ultimate answer is contingent on the manner in which the original question is postulated. Using an array of primary source data, gleaned largely from plantation records, census material, personal accounts, and other contemporary sources, scholars have created radically divergent interpretations of the question at hand. However, the overarching debate, in all of its many facets, can be reduced to a more simplistic dichotomy: (1) Was slavery as a tool for running a business sufficiently profitable for individual slave owners; and (2) Was slavery as an economic system profitable for the American South as a region?

From the point of view of historians, discussion of these central questions has been oriented around the pioneering work of two scholars: Ulrich B. Phillips and Kenneth M. Stampp. Writing more than four decades distant from one another, their answers to the question of slavery's profitability were diametrically opposed. Introducing his ideas in 1910s, at the height of the Jim Crow era, Phillips (1918) argued that southern slavery produced great wealth for some but was ultimately a moribund institution, leaving Dixie economically backward and subservient to the North. Stampp (1956), working at the dawn of the

American civil rights movement in the 1950s, conversely argued that slavery was not a dying system but was cruelly efficient and highly profitable.

For the past five decades, historians of varying calibers have taken up the mantle of one side or another of this debate. A number of recent scholars have even attempted to reconcile these two schools of thought. The work of Mark M. Smith (1998) is typical of this interpretation. In *Debating Slavery*, Smith argues that while slavery was profitable for individual slave owners, the system itself was detrimental to the South as a region.

However, arguments that deny the profitability of antebellum slavery, either from the perspective of individual business owners or of the South as a whole, are ultimately unsatisfactory. Indeed, slavery was a highly profitable business model that served to concentrate tremendous wealth in the hands of an elite class of southern planters. Further, slavery as an economic system proved to be a tool for the creation of new wealth and economic expansion for a young America marching westward.

In the America of today, devoid of any kind of sanctioned race-based slavery, it might be comforting to believe that the South's peculiar institution was lumbering toward its demise on the eve of the Civil War, as obsolete as vacuum tubes have become to the modern digital generation. Nothing could be further from the truth. On the eve of the Civil War, the South was experiencing tremendous economic growth with an increasing average per capita income outpacing that of the North.

Between 1850 and 1860, South Carolina, with the highest percentage of slaveholders in the nation, experienced an increase in real estate and personal property valuation of 90.15 percent, whereas New York's rate of increase was 70.63 percent. Both were home to major Atlantic ports and served as hubs of international trade, yet the advantage in personal wealth was clearly yielded to the southern state, despite its smaller population. In reality the slaves *themselves* represented wealth—they were not merely an economic burden to the owners' balance sheets, but a tangible asset—truly "human capital."

Thus the sad tragedy of slavery is epitomized. Rather than a derelict social construct of the ancient world, antebellum southern slavery was a pathway to individual and collective wealth. As such, the thesis offered by Phillips and others that slavery would have died of its own accord without the acrimony of Civil War is highly dubious. Indeed, American slavery is a case study in economic expediency triumphing over moral clarity.

The Nation the Slaves Made

The practice of African slavery was introduced into the region that would become the United States at Jamestown, Virginia, in 1619, a little over a decade after the colony's founding. The slaves' arrival was precipitated by the need for

an efficient, cost-effective workforce. In 1609 Jamestown settler John Rolfe (the future husband of Pocahontas) introduced a mild variety of tobacco from Bermuda, *Nicotiana tabacum*, to the fertile soils of Virginia. This forebear of the modern tobacco crop was palatable to the English market, unlike the harsh, rustic variety found growing in the region.

By 1612 Rolfe's plants were producing a paying harvest. At his estate, Varina Farms, Rolfe established the model of plantation agriculture that would endure for more than two centuries. Though he died in 1622 at the age of 37, Rolfe was enriched by the production of tobacco and his innovation placed the struggling Virginia colony on the road to economic prosperity. By 1620, the year after the introduction of the first slaves to Jamestown, the colony produced 119,000 pounds of tobacco—up from 2,500 pounds in 1616. From that point forward, slavery and cash-crop agriculture in the South were conjoined.

The transatlantic slave trade existed since approximately 1500. The first slaves to arrive in Jamestown in 1619, aboard a Dutch trading vessel, came as indentured servants—so-called bondsmen, who would eventually be granted their freedom. In 1654 John Casor became the first person to be legally recognized as a slave for life in the lands of the future United States. This precedent allowed for the exponential expansion of slavery throughout the American colonies. The legality of the system itself was gradually codified into English law thereafter.

Closely associated with the crops tobacco, sugarcane, and, most prominently cotton, slavery grew rapidly in the American colonies and, subsequently, in the newly minted United States. By the time transatlantic slave trade was outlawed in the United States in 1808, some 300,000 slaves had been directly imported, and an illegal trade in humans continued until the Civil War. Through "breeding," the domestic slave population also grew rapidly in America, in contrast to other regions where slaves barely reproduced themselves. American slave owners came to see slaves as a form of wealth and thus enhanced their investment by engaging in the internal slave trade. As such there were some 4 million slaves in the United States in 1860; a startling 12.8 percent of the total population, despite the fact that slavery was largely outlawed in the North by 1800. On the eve of the Civil War, the average price of a slave was $500, a record high.

Slavery as it existed in the American South came to fruition subsequent to Eli Whitney's patenting of the cotton gin in 1794. If an argument could be made for the moribund nature of slavery in America, then it must be applied to the period before the introduction of Whitney's machine. As a case in point, Thomas Jefferson, a Virginia slaveholder and architect of the Declaration of Independence, included a condemnation of slavery in his early draft of the Declaration, referring to it as a "cruel war raged against human nature itself." Though the Continental Congress struck this language from the document, such an admission by a leading Virginia planter is remarkable—and representative of the fact that slavery was facing a decline in the latter half of the 18th century.

Following in the wake of Whitney's invention, slavery experienced a rapid resurgence. The cotton gin allowed for the efficient, economical separation of cotton fibers from seedpods. As such, cotton could be produced cheaply using unskilled labor working in the fields and operating the gins.

This allowed for the mass production of cotton in an economy of scale. It also created the need for more land and the expansion of the plantation system into unsettled territories. The American slave system became the most immediate and profitable method of expanding this incipient cotton industry into the Deep South and, eventually, Texas.

The institution of slavery as it unfolded across the southland did so in the compacted space of a single human lifespan—then just a little over 60 years. Its arrival in the Mexican territory of Texas in 1821, where slavery was ostensibly illegal, proved to be a contributing factor to the independence and annexation of the Lone Star republic. Texas provides for an interesting case study of the expansion of slavery as a successful economic system. Southerners flowing into Texas established cotton culture in the eastern third of the future state, building tremendous fortunes in the space of only a few short years.

Historian Randolph B. Campbell (1989) points out in his *An Empire for Slavery* that though Texas was a region in which the South's peculiar institution was introduced late, it was adopted on a scale comparable to that found throughout Dixie. Indeed the expansion of cash crop agriculture based on cotton was a prime motivator in the settlement of the region of Texas. The threat to slavery posed by the centralized Mexican government further proved to be a contributing factor in Texans' push for independence.

Rather than a decadent institution, antebellum southern slavery was expansive and economically vigorous in the years leading up to the Civil War. Following the invention of the cotton gin, it became a brutal tool for opening up southern real estate to the forces of agricultural exploitation. Cotton became king of the Old South because as a cash crop, efficiently produced, it created a pathway for tremendous wealth.

Slavery by the Numbers

Though nearly 150 years has passed since the start of the Civil War, a good deal of information remains that will allow us to consider both the micro- and macro-economics of the slave system. Plantation records, census data, newspaper advertisements, personal accounts, and surviving slave narratives all contain information that sheds greater insight on this question. However, it must be understood that this is not a modern line of inquiry. Indeed, slavery's profitability was a hotly debated issue well before the coming of emancipation.

Antebellum abolitionists frequently argued that slavery was a dying, unprofitable institution. They reasoned that the master–slave relationship had degraded

Slaves picking cotton on a plantation. (North Wind Picture Archives)

Abolitionism

Though slavery was abolished in Sweden as early as the 14th century, for most of the world, abolitionist movements dated to the 18th century's burgeoning concerns with human rights and liberty. In many cases, these movements were tied to religious groups—the Quakers in both Britain and the United States, and various evangelical and Methodist groups in the United States. American protests against slavery had begun as early as the 17th century in colonies that did not depend on it economically, and by the turn of the 19th century, most northern states had abolished it on their own devices. The American colony of Liberia, on the west coast of Africa, was founded as a destination for freed slaves.

---

southern moral standards and character, leading to the corruption of the entire region. The South, they proposed, had fallen into economic backwardness precisely because of the peculiar institution.

As such, abolitionists made the case that the institution as a whole was an economic failure. Abolitionists could not attack the profitability of slavery on the part of individual plantation owners, however. There was no specific moral prohibition on profit taking in the antebellum world and as such it was not seen as unsavory. Condemning the planter class for personal enrichment was an argument

that would have made little sense, thus abolitionists sought to condemn slavery as economically detrimental to the region as a whole.

Regardless, many of those who condemned slavery as a corrupt and backward system were forced to admit its profitability for individual slaveholders. For instance, Frederick Law Olmstead (1856), the renowned landscape architect and traveling correspondent, reported firsthand in his *A Journey in the Seaboard Slave States* the devastating effects of slavery on the moral clime of the Old South. He condemned the institution as retrograde and inefficient, yet he was forced to admit its tremendous profitability for southern planters.

Thus the debate has arrived on the doorstep of modern historians. What remains is a confusing litany of propaganda from both sides of the antebellum dispute, drawn largely from the same sets of source data—or, for that matter, no data at all. A puzzling conundrum exists for those who would argue the unprofitability of slavery: its dogged persistence. The capitalist model insists that inefficient, unprofitable business practices will neither endure nor grow. Yet slavery did both.

Once again, in considering the question of slavery's profitability we are essentially bound by the rules of interpretation set by earlier scholars: (1) Was slavery as a tool for running a business sufficiently profitable for individual slave owners; and (2) Was slavery as an economic system profitable for the American South as a region? Let us take these questions in the order they are given and examine "the numbers"—the financial data that come to us from the historic record. This is a field many times plowed, and ample information exists from which to draw conclusions.

The Profitability of Slavery as a Business

As has been stated, the first historian to make a substantive inquiry into this question was Phillips. It must be noted that his work, today, remains largely discredited, not so much for his economic position, but for his overtly paternalistic and racist view of the slave system. Phillips's essential argument was that slavery, though moribund, was basically benevolent, elevating what Phillips considered to be the degraded nature of African Americans. For Phillips, slavery was a form of social control, a means of protecting whites from the so-called ravages of free blacks. As a result by and large of the odiousness of the racist doctrines expressed in Phillips's work, the man's publications have few defenders today.

Phillips's interpretation rapidly became the order of the day for a nation enmeshed in Jim Crow—the de jure system of apartheid that emerged following slavery's demise. However, in the following decades, a new generation of revisionist historians began to take a closer look at the profitability question. In 1942 Thomas P. Govan became one of the earliest scholars to rebuff Phillips and those writing in his vein. Govan used financial and population data from

census records coupled with extant financial records from plantations to conclude that slavery was a highly profitable business model.

In doing so, Govan took a comprehensive view of the notion of profitability. He argued that in calculating the profitability of slavery *as a business*, earlier historians had failed to understand the concept, insisting that a relationship existed between a given planter's profit figure and that planter's investment, when it did not. In short, plantation expenses and the lifestyle of those who resided on the plantation were intertwined.

Slavery provided the plantation owner "value-added" profitability in terms of labor: slaves ostensibly purchased to work in the fields could be put to work in other endeavors, such as clearing land, construction projects, household chores, maintenance tasks, and other jobs that in a wage-based business model would require additional expenditure. Further, the planter's way of life—his clothes, food, household, animals, entertainments, books, and virtually all other expenditures—was inseparable from the operating costs of running the plantation. The equivalent analogy would be a factory owner who lived within the walls of his plant and purchased all of his personal comforts on the balance sheets of the corporation. In this way, even plantations that turned only a small visible "profit" could provide the plantation owner with a tremendously rich lifestyle.

Govan's most startling example of this business model comes from the intact records of the plantation Hopeton from Glynn County, Georgia. Run by the Couper family from before the War of 1812 until the 1850s, the plantation's records provide consistent evidence of profitability in a continual manner throughout the life of the business. In this calculation, the slave workforce was not a debit, but an asset. During lean years, slaves were hired out or sold to augment profitability, a practice that could not have taken place with a hired workforce. Without the pressures of the traditional labor market (e.g., the inflationary demand of wages), plantation owners could continue production even in difficult years.

Further, the expenses of a plantation were all encompassing. The lifestyle of the plantation—regardless of extravagance—was deducted from the expense of the plantation itself, not from the owner's personal profit. When the pioneering geologist Charles Lyell visited Hopeton in 1845—during a period of decreased profitability throughout the South—he found the Coupers living in lavish style, with a fine estate, a well-stocked library of expensive volumes, and an enslaved workforce of 500, able to cater to any demand the plantation master might have.

With this in mind, the question of profitability must extend beyond itemizations of net and gross on a balance sheet. In short, Govan argued, slavery generated tremendous wealth for plantation owners in terms of the capitalized value of their farms and their slave workforce. The plantation was more than just an engine of production; it served as a direct means of providing the master with an often-extravagant lifestyle.

Stampp expanded on the work of Govan and others in 1956 with the publication of his seminal work, *The Peculiar Institution*. Though Stampp took a holistic view of the institution of slavery as a system of maintaining social order, as an economic system, and as a business model for individual planters, his conclusions concerning profitability were unequivocal. In short, after analyzing plantation records, sales and pricing data, and census data, Stampp found that slavery delivered for its practitioners an exceptional profit above and beyond the cost of expenditure. Further, Stampp explained, this profitability was not merely the result of the sale of cash crops, but also the result of the hiring out and sale of slaves.

Slaves themselves constituted appreciable wealth, unlike the expenditures incurred by a typical labor force. In a chapter dedicated to "Slavemongering," Stampp explained the use of slaves as a means of augmenting profit. In addition to being hired out or being "bred" for sale, they also served as a form of capital that could be liquidated to clear other debt. In the 1840s, during a period of poor markets in the South, slaveholders routinely reduced their workforces not merely as a means of reducing expenses, but also as a method for generating liquid capital that could then be used to service debt. Such a practice would be impossible with a traditional labor market and thus allowed plantation owners to hold back the rising tide of foreclosure, carrying them to the next economic upturn.

Stampp found slavery to be a profitable enterprise for slave owners, regardless of the size of their holdings. Even as slave prices increased throughout the 1850s, small-scale farmers used slavery to their economic advantage. Stampp indicated that the presence of a few slaves on a farm, able to perform work for the cost of food and clothing (which slaves grew and made themselves), meant the difference between subsistence and bankruptcy. For large farmers, slavery was the key to prosperity. Stampp articulated that antebellum observers, even antislavery thinkers such as Olmstead, were forced to admit that slavery-based agriculture provided a better return on investment than any analogous northern business strategy.

Following its introduction, Stampp's work was seen as controversial, overturning four decades of a deeply ingrained interpretive paradigm. Slavery was no longer allowed to be seen as a withering artifact at the time of its abolition, but rather a vigorous vehicle for personal wealth with a far-from-certain demise. In Stampp's wake came additional refinements of his ideas. Stampp, who was not an economist, lacked many of the intellectual tools to verify his suppositions. This verification came chiefly from the partnership between Robert W. Fogel and Stanley Engerman.

Writing from the perspective of economists, these authors confirmed Stampp's understanding of slavery as a profitable business model. Though earlier authors such as Alfred H. Conrad and John R. Meyer (1958) had attempted to do much the same, Fogel and Engerman's *Time on the Cross* (1974) brought

together a litany of quantitative data, gleaned from myriad antebellum sources, to affirm the profitability of slavery. Their argument was grounded in both micro- and macroeconomic principles. They confirmed the conclusion reached by Stampp that, in fact, slavery was a profitable business model, that economic growth experienced in the South in the years leading up to the Civil War was greater than that of the North, and that only the most inefficient farmers and planters failed to extract a profit from their chattel.

Fogel and Engerman envisioned southern slavery as a highly specialized labor system, rapidly perfected in the years following the development of the cotton gin. Subsequent to the firm establishment of slavery in the Deep South, slave prices continued to climb, creating a profitable market in the domestic slave trade. If speculating planters were pessimistic about the future of slavery, as the likes of Phillips argued, then their willingness to engage in the sale and purchase of slaves, at inflated prices, belied this fact.

Fogel and Engerman's work created a firestorm among historians that drifted into the popular press. Their stance on the issue of slavery's profitability quickly overshadowed that of Stampp's. Over the past three decades, much ink has been spilled in an attempt to refute Fogel and Engerman's thesis. Fogel, himself, refined his ideas in an additional book in 1988 that backed away from many of the lesser-substantiated claims made in *Time on the Cross*.

Yet today most economic historians concede that the claim that slavery was a profitable business model has been amply verified. In 1995 Robert Whaples conducted a survey of the Economic History Association and found that nearly 100 percent of the membership's economists and historians agreed with Fogel and Engerman's profitability thesis. By a further high percentage, a majority of economists and historians agreed that slavery served as an efficient business model for farmers and planters. This is evidence of the sea change Stampp's original ideas have brought about—indeed, Phillips has been utterly rebuffed. More than a means of social control, the odious system of slavery stood as a pathway to personal economic gain, making its voluntary abolition all the more unlikely.

The Profitability of Slavery as an Economic System

The notion that slavery was unprofitable and, indeed, detrimental to the economic well-being of the South finds its origins in many quarters. Antebellum observers commented on Dixie's "backwardness" in comparison to the North. It lacked the kind of industrial and infrastructure development more common above the Mason-Dixon Line. Southerners themselves were tied to the land, isolated in rural communities with little exposure to the educational and spiritual inducements found in the urbanized regions of the country. Abolitionists ascribed this "backwardness" to the institution of slavery—arguing a direct connection, where none exists. Following in their vein, more recent historians, most

notably Eugene Genovese (1974) in *Roll, Jordan, Roll*, have made a similar argument.

Indeed, the South's "backwardness," if we are forced to call it that, was not so much a product of slavery as it was in building an entire economy around the system of cash-crop agriculture. In reality, the backwardness found in the antebellum South is conducive to agricultural-based economies the world over, both before and after the abolition of slavery. The South forced itself into the model of a colonial economy.

Large plantations were closed systems, producing much of their own food-stuffs, clothing, and other goods that would normally be purchased as manufactures. The economic system of agriculture, then, as today, was heavily reliant on credit. Credit had to be extended to farmers and planters in the absence of liquid capital, only to be repaid, with interest, at harvest time. Good production years produced repayment and allowed the return of profit, while bad years might leave little profit or, worse, the inability to repay debt. This is a familiar tale common to farmers and planters in the modern world—a world devoid of slavery.

In calling the antebellum South "backward," historians indicate a view of that region as somehow separate and distinct from the total economy of the United States. It is evident that the South lacked the industrial development to persevere as an independent nation, as the Civil War aptly demonstrated. But to measure the antebellum South by the standards of its northern counterpart and ascribe to it the label of "backward" is analogous to the modern observer referring to the present-day Midwest as somehow more "backward" than corresponding regions of the United States.

In truth, much of the Midwest lacks industrial development, large-scale urban populations, and major cultural and trading centers. As a region primarily based on an agricultural economy, the Midwest of today, like the South of old, is a colonial economy—incapable of economic subsistence without external capital and markets. Yet that system exists—and persists—devoid of slavery. The key difference, if we continue with this analogy, is that in the decades approaching the onset of the Civil War, the South actually maintained *higher per capita income growth* for all citizens, both slaveholding and nonslaveholding, than their northern counterparts.

Eugene Genovese (1965), in *The Political Economy of Slavery*, has made the argument that the South's rapid economic growth was the result of a boom in global demand for cotton beginning in 1820. Somehow, this observation is intended to counterbalance the notion of slavery as a path to wealth. Rather, slavery proved to be the most efficient method for southern planters to fulfill this increased demand. Fogel and colleagues themselves, in *Without Consent or Contract*, found in refining their ideas that the prewar cotton boom accounted for only about one-fifth of the South's increase in per capita income. Indeed, the capitalized wealth and income from the slave market, itself, cannot be left

out of the equation. Southerners' wealth and prosperity was not simply achieved on the backs of slave labor; the slaves themselves constituted much of that wealth.

Authors Jeffrey Hummel (1996) in *Emancipating Slaves, Enslaving Free Men* and Thomas J. DiLorenzo (2002) in *The Real Lincoln* have revitalized the notion of slavery's moribund nature by arguing that the slave system had to be propped up through the passage of the Fugitive Slave Act. Following a line of reasoning first offered by abolitionist William Lloyd Garrison (who advocated *northern* secession in order to achieve this end), DiLorenzo argues that the nullification of the Fugitive Slave Act would have collapsed the slave market through inflation and destroyed slavery as an economic system. DiLorenzo makes the claim—similar to Phillips—that because slavery was so fragile, the Civil War, which he interprets as an act of northern aggression, was utterly unnecessary and the fault of Abraham Lincoln's near tyrannical leadership.

Such a proposal conveniently ignores the case of Texas. As George R. Woolfolk (1956) in "Planter Capitalism and Slavery: The Labor Thesis" and Randolph B. Campbell (1989) in *An Empire for Slavery* have aptly demonstrated, slavery built tremendous economic prosperity for Texas within the span of only 40 years; this despite the fact that the system was technically illegal there from 1821 until Texas independence in 1836. Further, the effect of the Fugitive Slave Act on the economy of Texas was essentially negated. Texas slaves did not need to escape north of the Mason-Dixon Line to achieve their freedom. Slavery was illegal in Mexico, and the Mexican nation, under its multifarious guises throughout the early 19th century, remained hostile toward the South's peculiar institution.

Texan slaves needed only to make their way to the other side of the Rio Grande as a means of achieving freedom. With whole swaths of the population in the Texas hill country and portions of the state's southwestern region openly hostile to the forces of slavery and disunion, such escapes occurred frequently. Yet slavery not only persisted in Texas, despite Garrison's argument, it prospered. Census data show the increase in personal and real property values in Texas between 1850 and 1860, fueled by slavery, to be a shocking 592.44 percent, the third highest rate of growth in the nation. In this light, one sees not a slow fading of a dying system, but a wave of economic prosperity generated by slavery that would have surely flowed across the nation, perhaps eventually to the Pacific Ocean.

Conclusion

The notion that slavery was an unprofitable system for southerners and the southern economy lacks foundation. Indeed evidence indicates that slavery served as a pathway for tremendous personal wealth and an enhanced lifestyle on the part of slaveholders, while at the same time providing a rapidly increasing level or per capita income and property valuation throughout the South. The

statement that the South was somehow "backward" as a result of slavery is dependent on one's definition of "backwardness." In terms of industrial development and urban growth, the antebellum South was not on par with the North, but in terms of personal wealth and property valuation, the South fared comparably well and thus lacked an economic incentive to change.

The arguments over the economic potential of slavery should in no way be tied to the morality of that despicable institution. Historians and economists who argue for the profitability of slavery are not seeking to be apologists for the South's peculiar institution. To the contrary, this argument supports the notion expressed by many abolitionists that slavery could not be stripped from the South easily. Those who argue the moribund nature of slavery fail to grasp that southerners—even those who were not slaveholders—were willing to fight and die in a war to preserve their way of life and the economic system on which it was built. This act of desperation did not emerge from a sense that slavery was dying, but rather from a feeling that the basis of southern prosperity was being torn away.

References and Further Reading

Campbell, Randolph B. *An Empire for Slavery: The Peculiar Institution in Texas, 1821–1865*. Baton Rouge: Louisiana State University Press, 1989.

Conrad, Alfred H., and John R. Meyer. "The Economics of Slavery in the Ante Bellum South." *Journal of Political Economy* 66, 1 (1958): 95–130.

Deyle, Steven. "The Irony of Liberty: Origins of the Domestic Slave Trade." *Journal of the Early Republic* 12, 1 (Spring 1992): 37–62.

DiLorenzo, Thomas J. *The Real Lincoln: A New Look at Abraham Lincoln, His Agenda, and an Unnecessary War*. Roseville, CA: Prima Publishing, 2002.

Egerton, Douglas R. "Markets without a Market Revolution: Southern Planters and Capitalism." *Journal of the Early Republic* 16, 2 (Special Issue on Capitalism in the Early Republic, Summer 1996): 207–21.

Fogel, Robert William and Stanley L. Engerman. *Time on the Cross: The Economics of American Negro Slavery*. New York: Little, Brown, and Co., 1974.

Fogel, Robert William, Ralph A. Galantine, and Richard L. Manning. *Without Consent or Contract: The Rise and Fall of American Slavery, Evidence and Methods*. New York: Norton, 1988.

Foley, Neil. *The White Scourge: Mexicans, Blacks, and Poor Whites in Texas Cotton Culture*. Berkeley: University of California Press, 1997.

Freudenberger, Herman, and Jonathan B. Pritchett. "The Domestic United States Slave Trade: New Evidence." *Journal of Interdisciplinary History* 21, 3 (Winter 1991): 447–77.

Genovese, Eugene D. *The Political Economy of Slavery: Studies in the Economy and Society of the Slave South*. New York: Pantheon, 1965.

Genovese, Eugene D. *Roll, Jordan, Roll: The World the Slaves Made*. New York: Pantheon, 1974.

Govan, Thomas P. "Was Plantation Slavery Profitable?" *Journal of Southern History* 8, 4 (November 1942): 513–35.

Hummel, Jeffrey Rogers. *Emancipating Slaves, Enslaving Free Men: A History of the American Civil War*. Peru, IL: Open Court, 1996.

Olmstead, Frederick Law. *A Journey in the Seaboard Slave States; With Remarks on Their Economy*. New York: Dix and Edwards, 1856.

Phillips, Ulrich Bonnell. *American Negro Slavery: A Survey of the Supply, Employment and Control of Negro Labor as Determined by the Plantation Regime*. New York: Appleton, 1918.

Scarborough, William K. "The Cotton Gin and Its Bittersweet Harvest." *Journal of American History* 81, 3 (The Practice of American History: A Special Issue, December 1994): 1238–43.

Secretary of the Interior. *Statistics of the United States, (Including Mortality, Property, &c.,) in 1860: Compiled from the Original Returns and Being the Final Exhibit of the Eighth Census, Under the Direction of the Secretary of the Interior*. Washington, DC: Government Printing Office, 1866.

Smith, Mark M. *Debating Slavery: Economy and Society in the Antebellum American South*. New York: Cambridge University Press, 1998.

Stampp, Kenneth M. *The Peculiar Institution: Slavery in the Ante-Bellum South*. New York: Knopf, 1956.

Whaples, Robert. "Where Is There Consensus among American Economic Historians?" *Journal of Economic History* 55 (1995): 137–47.

Woodman, Harold D. "The Profitability of Slavery: A Historical Perennial." *Journal of Southern History* 29, 3 (August 1963): 303–25.

Woolfolk, George R. "Planter Capitalism and Slavery: The Labor Thesis." *Journal of Negro History* 41, 2 (April 1956): 103–16.

15

Lincoln maneuvered the South into firing the first shot at Fort Sumter.

PRO   Rolando Avila
CON   Lee Oberman

PRO

The United States of America was founded with the principle of compromise as a chief cornerstone, but Abraham Lincoln refused to avert war through compromise. Although the right to secession was a highly controversial and debated issue, Lincoln refused to accept secession on any grounds. In fact, he threatened the use of force if he perceived any threat to the integrity of the Union. More patience and a policy of peace might have allowed cooler heads to prevail. The Union could possibly have been preserved without such a costly war. However, Lincoln's unwillingness to honor the long-standing American political tradition of compromise and his war policy maneuvered the South into firing the first shot at Fort Sumter, which started the war.

Compromise had been a part of American government from the beginning. In 1787 delegates to the Constitutional Convention debated over key issues related to representation, and their compromises shaped the framework of the federal government. Representatives from smaller states feared that they would be at a political disadvantage to larger states. So a compromise was reached that called for proportional representation in the House of Representatives based on population and equal representation of each state in the Senate. Also, representatives from the South protested that slaveholding put them at a disadvantage, since the North comprised many more eligible voters. Consequently, the Three-Fifths Compromise allowed, for the purpose of representation, five slaves to count the same as three free people.

The delegate to the Constitutional Convention also wrestled with the tension of states' rights verses federal rights. Some delegates feared that if the federal government were given too much power, states' rights would be trampled on. Others feared that if states were given more power than the federal government, this would lead to the creation of separate nation–states instead of one united nation. In order to secure the necessary votes for ratification of the U.S. Constitution, their solution was to create a system in which power was ambiguously shared. Since then, advocates of federal and states' rights were able to turn to

333

the Constitution and find support for their respective views on the division of powers.

Later, as sectional problems arose, political compromises continued to help smooth out differences. In 1820, for example, the U.S. Congress reached a compromise that allowed Missouri to be admitted to the Union as a slave state and Maine as a free state. But more important, the compromise served to balance political power in the Union, which as a result of the compromise now consisted of 12 free and 12 slave states. Another notable compromise in 1850 sought to please both the North and the South. It admitted California to the Union as a free state, abolished the slave trade in Washington, D.C., created the territories of New Mexico and Utah without restrictions on slavery, and strengthened the fugitive slave laws.

In 1860 a protest wave occurred in the South following Lincoln's election, and more states threatened to secede if Lincoln took office because many Southerners believed Lincoln would abolish slavery. Senator John Crittenden introduced the Crittenden Compromise in the U.S. Senate in an effort to address some of the unrest that was threatening to tear the nation apart. Similar to the Compromise of 1850, the Crittenden Compromise, which was a collection of six constitutional amendments, sought to limit slavery without abolishing it. Although it did not attempt to settle the slavery question for good, it did give hope of a temporary peace as all previous compromises had done. Many representatives from both North and South liked it and began seeking support for it. However, President-elect Lincoln kept silent. Some historians have contended that if Lincoln had supported the compromise, it would have gained the needed support and would have surely passed. However, without Lincoln's support, the compromise failed and the spirit of compromise faded from Congress.

Critics who have claimed that, at best, the compromise would have only postponed war have been somewhat shortsighted, for in all human history peace has been temporary and enjoyed only in installments. Also, those that have claimed that it was better to face the issues and go to war in 1861 than to have put it off have concluded this by looking at history as only a thing of the abstract past. Would these same judges of past virtue venture to say the same about breaking the peace today for political disagreements? Or would they much more likely be inclined to seek out solutions for peace, however temporary? Over the years, the estimates of the casualties of the American Civil War, not including the high civilian casualties, have varied, but the most accepted figures of the military death toll have ranged between 165,000 and 170,000. Furthermore, hundreds of thousands of soldiers were wounded. Many wounded soldiers were affected for the rest of their lives by the loss of limbs and other physical and mental injuries. Soldiers' families were affected by the losses on the battlefields as well. Since in 1860 most Southerners did not want disunion and most Northerners did not want to press the slavery issue, the high cost of the failure to reach a compromise became ever more apparent and tragic.

Abraham Lincoln on Slavery

Although never a friend of slavery, Abraham Lincoln was certainly willing to tolerate its existence if it would have saved the Union. His ambivalence toward slavery itself can be seen in his speech regarding the repeal of the Missouri Compromise in 1854:

> Before proceeding, let me say I think I have no prejudice against the Southern people. They are just what we would be in their situation. If slavery did not now exist amongst them, they would not introduce it. If it did now exist amongst us, we should not instantly give it up. This I believe of the masses north and south. Doubtless there are individuals, on both sides, who would not hold slaves under any circumstances; and others who would gladly introduce slavery anew, if it were out of existence. We know that some southern men do free their slaves, go north, and become tip-top abolitionists; while some northern ones go south, and become most cruel slave-masters.
>
> When southern people tell us they are no more responsible for the origin of slavery, than we; I acknowledge the fact. When it is said that the institution exists; and that it is very difficult to get rid of it, in any satisfactory way, I can understand and appreciate the saying. I surely will not blame them for not doing what I should not know how to do myself. If all earthly power were given me, I should not know what to do, as to the existing institution. My first impulse would be to free all the slaves, and send them to Liberia—to their own native land. But a moment's reflection would convince me, that whatever of high hope, (as I think there is) there may be in this, in the long run, its sudden execution is impossible. If they were all landed there in a day, they would all perish in the next ten days; and there are not surplus shipping and surplus money enough in the world to carry them there in many times ten days. What then? Free them all, and keep them among us as underlings? Is it quite certain that this betters their condition? I think I would not hold one in slavery, at any rate; yet the point is not clear enough for me to denounce people upon. What next? Free them, and make them politically and socially, our equals? My own feelings will not admit of this; and if mine would, we well know that those of the great mass of white people will not. Whether this feeling accords with justice and sound judgment, is not the sole question, if indeed, it is any part of it. A universal feeling, whether well or ill-founded, can not be safely disregarded. We can not, then, make them equals. It does seem to me that systems of gradual emancipation might be adopted; but for their tardiness in this, I will not undertake to judge our brethren of the south.
>
> When they remind us of their constitutional rights, I acknowledge them, not grudgingly, but fully, and fairly; and I would give them any legislation for the reclaiming of their fugitives, which should not, in its stringency, be more likely to carry a free man into slavery, than our ordinary criminal laws are to hang an innocent one.

Source: Abraham Lincoln, Speech at Peoria, Illinois, October 16, 1854.

Those who have contended that going to war was good because it netted slaves their freedom have failed to consider that no one in 1861 knew what the future held in store. Furthermore, Lincoln's reasons for rejecting secession and going to war had nothing to do with slavery. Lincoln's reasons were economic. The North was dependent on the South for raw materials (such as cotton for textiles). Lincoln's reasons were political. Lincoln believed that if any state were allowed to leave the Union, it would establish a precedent so dangerous that it would threaten the safety of the Union in the future. In Lincoln's opinion, his main objective was to preserve the Union, and if keeping all the slaves bound accomplished this goal, he would do it.

Some historians have gone so far as to claim that the Civil War was not necessary in order to bring about the end of American slavery. Their proof for this was that slavery was dying out all over the world. North American states had done away with slavery many years before the war, and slavery was on its way out in many other areas of the Union including Missouri, East Tennessee, and Western Virginia. It would have taken several more years for emancipation to occur without the war, but emancipation was on its way. It would eventually have happened with or without the war.

Abraham Lincoln won the presidential election of 1860. However, the election constituted a breakdown of the fragile political party system. Although Lincoln (Northern Republican) won both the popular votes (1,866,452) and electoral votes (180), the South was greatly underrepresented. Stephen Douglas (Southern Democrat) won almost as many popular votes (1,376,957) as Lincoln, but fell astonishingly short on electoral votes (12). John C. Breckinridge (Southern Democrat) received fewer popular votes (849,781) than Douglas, but beat him with electoral votes (72). John Bell (Northern Constitutional Union) won fewer popular votes than any other candidate (588,879), but he also earned more electoral votes (39) than Douglas. The election showed that there was an obvious difference in what people wanted and what the political system votes reflected. To make matters worse, Lincoln did not receive a single popular vote in most of the states in the Deep South. In fact, he did not even appear on the ballots of some southern states. Although 10 of the 11 future Confederate states did not cast a single popular vote for Lincoln, he still became president. Many frustrated Southerners felt that Lincoln was not their president. They believed they had not elected him and he did not represent their interests and values. Although President Lincoln tried to assure the South that he had no intentions of abolishing slavery, they did not believe him. Consequently, the presidential election of 1860 led to the secession crisis. By February 1, 1861, South Carolina, Mississippi, Florida, Alabama, Georgia, Louisiana, and Texas had seceded from the Union and formed the Confederate States of America.

The threat of secession was nothing new in 1860. It rested on the constitutional doctrine of state sovereignty. According to this view, states entered the Union voluntarily by state conventions, and they could leave the Union in the same way.

During the War of 1812, New Englanders suffering from economic sanctions were the first to threaten the federal government with secession unless they were given some relief. Also, in 1828, South Carolina planters protested the Tariff of 1828 by threatening to nullify, or choose not to obey, federal laws that hurt its interests. The threat of nullification led to a nine-day debate in Congress between Senator Robert Y. Hayne of South Carolina and Senator Daniel Webster of Massachusetts. Hayne insisted that the federal government had no right to hurt South Carolina, which was a sovereign state. Webster, on the other hand, believed that if a state were allowed to declare an act of the federal government null and void within state limits, this, in essence, would make the state completely autonomous rather than part of a larger and inseparable Union. President Andrew Jackson agreed with Webster and threatened to send troops to enforce federal law, but the nullification crisis was mostly subdued by a congressional compromise tariff in 1833.

A variation of the right to secession stated that only Texas, which had been a sovereign nation before it was annexed into the United States, and the original 13 states that signed the Declaration of Independence in 1776 had the right to secede. On the other hand, none of the other states in the Union could claim this right, because the federal government had created them all from government-owned lands. However, Lincoln denied that any state had the right to secede. During the secession crisis, Lincoln was unmovable. Inseparable Union was an obsession with him.

In his first Inaugural Address, on March 4, 1861, he stated that he understood the Constitution to make the Union of the states perpetual. In his view, no government had ever created the means for its own disintegration. He stated that disunion would contradict the Constitution that he was bound to uphold. Consequently, even during the secession crisis, he stated that the Union must remain unbroken. Furthermore, he did not recognize the Confederate States of America (CSA), but, rather, he often referred to them as rebels or Union states that were temporarily in rebellion.

Abraham Lincoln in 1864. (Library of Congress)

As soon as Lincoln took office, he began to feel the great political pressures of his new job. On one side was the radical faction that wanted to teach the South a lesson through as much force as possible. They believed that the preservation of the

Union was the most important issue at hand. On the other extreme was the peace faction that wanted no destruction of the South at all. In fact, many peace proponents felt it was preferable to allow the South to secede peacefully rather than have war. Although more of a moderate himself, Lincoln tended to side with the radical faction. He did not want to give up any federal fort. As far back as December 22, 1860, when he was still president-elect, he expressed this strongly held view in a letter to Major David Hunter. Lincoln wrote that, in his view, all federal forts should be held, and if they fell they should be retaken. A couple of days later Lincoln heard unconfirmed rumors (rumors that he later found out to be false) that President James Buchanan was considering ordering federal forts in South Carolina to surrender. Lincoln sprang into action and wrote to Senator Lyman Trumbull, instructing him to announce publicly that, if the rumors were true, Lincoln would retake the forts as soon as he was sworn in.

Lincoln never strayed from his conviction to hold federal forts. And once he was in office, he did not want to let Fort Sumter go under any circumstances, because he feared that the loss of the fort would hurt the authority and prestige of his new administration and party. A letter from Major Robert Anderson, the commander of Fort Sumter, informed Lincoln that, unless supplies were sent soon, he would be forced to abandon the fort. Therefore, there was great urgency for Lincoln to devise a plan, and he began to consider his options.

If he acted aggressively, as many men in his party wanted, he ran the risk of losing more states to the Confederate cause. On the other hand, if he failed to act, he would alienate the radical republicans, which made up the majority of his party. If he ordered the fort to be evacuated or if he let supplies run out so that Anderson would be forced to evacuate, then he ran the risk of appearing weak before the Confederacy and the world, which might improve the Confederacy's chances for foreign aid or even recognition. He struggled with the complexities of the problem that he knew he alone would have to make a decision on.

Lincoln sought advice from his cabinet and from army commander General Winfield Scott. All of the cabinet members, except Montgomery Blair, agreed that sending supplies to the fort would be a mistake, because it would undoubtedly lead to war. In Scott's opinion, no military force could secure the fort, and he recommended evacuation. However, Blair continued to push for a military operation and recommended Captain Gustavus V. Fox to lead several small supply boats under the cover of darkness. Although his cabinet and Scott advised against it, Lincoln refused to budge from his conviction, which he had held even before he took office, that federal forts must be held or retaken if necessary.

So Lincoln sent Fox to Charleston Harbor acting as his agent under the guise of a peace mission. Fox's real goal, however, was to get familiar with the geography of the harbor and to begin to devise a plan for the best way to enter the harbor with his supply boats. Next, Lincoln sent Ward Hill Lamon, a longtime friend and personal bodyguard, to Charleston so that he could get a sense

of the sentiment in the area. Governor Francis W. Pickens informed Lamon that any sign of a U.S. warship would draw Confederate fire. However, he would allow a steamer to enter the harbor peacefully to transport the men out of the fort. Lamon promised Pickens that he would soon return with an evacuation order from Lincoln.

During this time, Secretary of State William H. Seward, Lincoln's closest adviser, kept trying to convince Lincoln to evacuate Fort Sumter before it led to war. Seward suggested that Lincoln should hold on to another less contested federal fort that could, just as well, stand as a symbol of federal authority. But Lincoln would not have it. By this point, Fort Sumter had become such an important symbol that it was too late for him to call retreat. However, Seward felt so strongly in his case that he met with Confederate commissioners and informed them that an evacuation would most likely occur soon. Somehow Seward's plan to evacuate the fort leaked out and many newspapers in New York, Chicago, Philadelphia, and Cincinnati printed the story.

With time running out for Fort Sumter, letter writers and visitors to the White House continued to exert pressure on Lincoln to remain loyal to his party by holding and supplying the fort. However, Lincoln considered the horrifying consequences of a military defeat, which Scott had predicted. How could he take the risk of a doomed military operation? Lincoln also realized that any show of force by the North would alienate peace factions in both the Democrat and Republican parties. Furthermore, it would solidify support for the Confederate cause in the South.

But, what if the South fired on the fort first? Would this not convince the border states to stay in the Union and solidify support for the Union cause in the North? Finally, Lincoln saw a clear path within his political constraints and he devised a plan to force the South to fire on the fort. Lincoln would turn the tables on the Confederates. Lincoln sent word to Pickens that Lamon had no authority to speak for him. In other words, there would be no evacuation. Also, Lincoln refused to meet with the Confederate commissioners, which Seward had met with, for fear that this meeting would be construed as recognition of the Confederacy. As a consequence, the Confederate agents sent word to Jefferson Davis that their peace mission had failed.

Lincoln sent Fox to New York to assemble a fleet and begin preparations for the supply mission. On the same day, Lincoln, in spite of objections from his cabinet, wrote a notice to Pickens that he would be sending a supply mission that would not attack unless acted on. The notice restated what Lincoln had been ensuring the South all along: There would be no war unless the South brought it about. Robert S. Chew of the State Department arrived by train in Charleston to delivery Lincoln's message on the same day that the supply expedition departed New York on a three-day voyage to the fort. This gave Confederates ample time to prepare for the arrival of the fleet, and they began immediate preparations.

Why did Lincoln notify the Confederates? Didn't he know that they would take the notice as a direct challenge? From the point of view of the Confederates, secession was justified on moral, legal, and constitutional grounds, and seceded states had the right to take over public property found within its boarders. So, the moment South Carolina seceded from the Union on December 20, 1860, the forts in South Carolina belonged to the Confederacy and they began taking them. About one week after secession, the South Carolina militia seized Fort Pinckney and Fort Moultrie, and on January 2, 1861, the militia seized Fort Johnson. Fort Sumter was one of the few holdouts. During the waning days of the Buchanan administration, the steamer *Star of the West* had made an attempt to supply Fort Sumter, but Confederates were able to drive it away by firing cannons in its direction from the shore. To borrow a historian's analogy, what if Great Britain had kept a fort in Boston or New York Harbor after the American Revolution? Would Americans stand for such an insult and threat to their safety? In the same way, the Confederate government could not allow the federal fort to stand and at the same time claim its independence from the Union.

Lincoln knew what he was doing. He was a master politician. Lincoln made calculated moves like a chess master. Rather than rashly taken, his moves were premeditated and cunningly executed. He was able to put the South at a disadvantage. All the pieces were in place. It was now just a matter of time. Would the South let the supply ships accomplish their mission and face humiliation and a serious setback in their efforts to establish themselves as a sovereign nation? Lincoln had put their honor at stake, and they would be forced to fire on the fort.

On April 9 the Confederates intercepted Anderson's mail and discovered a letter to the War Department in which Anderson discussed Lincoln's plan to supply the fort by force. With this recent discovery, Confederate action became even more urgent. Confederates made one last attempt to avoid bloodshed. They gave Anderson one last opportunity to surrender and evacuate the fort. But, being under orders to stay in the fort and wait for supplies, Anderson declined the offer and, instead, only evacuated women and children. Confederates believed that the Union was exhibiting aggression by refusing to yield the fort. There was no way around it. Confederates had to fire on the fort, and they had to fire on it before the fleet arrived. Otherwise they would have to contend with both the fort and the fleet at the same time. Jefferson Davis ordered General Pierre Beauregard, Confederate commander in Charleston, to capture Fort Sumter at once.

On April 12 the Confederates opened fire on the fort. The continuous booming blasts from the cannons woke up Charleston residents who climbed on rooftops to get a better view of the attack. Although the bombardment of the fort lasted for 34 hours, U.S. warships stationed just outside of Charleston Harbor made no attempt to enter the harbor and render aid. In fact, not even the supply boats reached the fort. They were caught in a storm and were forced to land in other ports. Faced with a hopeless situation, Anderson lowered the U.S. flag and

raised a white flag to surrender. This military failure had achieved Lincoln's intended objective. By firing on the fort, Confederates had played right into his hands. Lincoln had maneuvered the South into firing the first shot, and, just as Lincoln had expected, news of the aggression unified the North against the rebels.

After Fort Sumter fell, Lincoln knew that he had to prepare for war. On April 15, he made a call for 75,000 three-month volunteers from the states to protect Washington. Many governors telegraphed Lincoln that they would easily meet and even exceed their quotas. Many Northerners were appalled by the fact that Confederates had fired on a fort that was flying the U.S. flag. To add insult to injury, many Northerners were angered by the fact that the former federal fort was now flying a Confederate flag instead of a U.S. flag. More than ever before, Fort Sumter became a symbol of southern treachery and northern patriotic fervor. The symbolism of the fort endured throughout the war. In fact, a ceremony was held at Fort Sumter five days after the end of the war. Anderson returned to Fort Sumter with the battle-torn U.S. flag that he had lowered during the surrender, and he raised it up again as a sign of the preservation of the Union and a celebration of the end of the war.

About one week before the Confederates fired on Fort Sumter, both the antiadministration and proadministration press accused Lincoln of forcing Southerners into a corner and leaving them no choice but to attack the fort. Editors of the Columbus, Ohio, *Crisis*, the *New York Times*, and the *Baltimore Sun* sounded the alarm. After their prophetic predictions came true, the Providence *Daily Post* immediately asked why the Civil War had begun and answered that Lincoln was to blame for the war, because he had put the safety of his party above the safety of the country.

Davis delivered a speech shortly after the events at Fort Sumter and described Lincoln's strategy during the crisis as a skillful "maneuver" to force the South to fire on the fort. The speech was later included in his *The Rise and Fall of the Confederate Government* (1881). Davis explained that the fort belonged to the Confederacy, which had been very patient with the North in an effort to avoid war, but Lincoln destroyed these peace efforts by sending a fleet to Fort Sumter. According to Davis, he could not allow the fort to be supplied, because it meant maintaining a threat to Charleston. So, according to Davis, the Confederate attack on the fort was simply an act of self-defense. To paraphrase an analogy that Davis included in order to better clarify his point, to allow the supply of Fort Sumter would have been like not batting away the hand of an aggressor until after he had had a chance to fire into your chest. Alexander Hamilton Stephens, vice president of the CSA, also accused Lincoln of the same manipulation in his *A Constitutional View of the Late War between the States* (1868). In his view, Lincoln was clearly the aggressor and was to blame for starting the war, because he had made force necessary by refusing to yield forts located in the Confederacy. Under the circumstances, Stephens insisted, all the Confederacy could do was respond in self-defense.

John Nicolay and John Hay, Lincoln's personal secretaries, praised Lincoln for essentially the same things that critics accused him of during the Fort Sumter crisis in their monumental 10 volume *Abraham Lincoln: A History* (1890). While many Southerners accused Lincoln of either being a bungler or a cunning villain, Nicolay and Hay proclaimed him a strategic genius. They were both very privileged insiders. As part of their secretarial duties, they were regularly in Lincoln's presence and overheard and discussed matters before the president. They read through and processed his incoming mail. In fact, they were the first biographers to get access to Lincoln's personal papers while writing their Lincoln biography. After Lincoln's death, they wrote that he maneuvered the South to fire the first shot, because he wanted to show the world that, not only was the North right, but also that the South was wrong. In the coming years, John Hay enjoyed an extremely successful political career.

Besides these high-profile individuals, many more historians agreed with and further elaborated on the maneuver thesis. Some accused Lincoln's attempt to provision the fort while at the same time informing the governor of his intentions as silent aggression intended to induce an active aggression from the South. Some of the most notable studies were the following. Charles W. Ramsdell's "Lincoln and Fort Sumter" in the *Journal of Southern History* (1937) was the first scholarly analysis of the event and has remained the starting point for researchers. According to Ramsdell, the Confederate government wanted peace and tried to avoid war by negotiating a peace settlement with Lincoln, but he refused to listen. John S. Tilley's *Lincoln Takes Command* (1941) was one of the first full-length studies to elaborate on the foundation set by Ramsdell. The book was published during World War II at a time when President Franklin D. Roosevelt was being accused of maneuvering the first shot at Pearl Harbor in order to garner American support to enter the war. In the chapter titled "Lincoln Got What He Wanted," Tilley accused Lincoln of aggressively pursuing war. David M. Potter's *Lincoln and His Party in the Secession Crisis* (1942), also published during World War II, accused Lincoln of being too closed-minded to compromise during the crisis. According to Potter, Lincoln could have averted war if he had been willing to negotiate with the South. Some recent studies that have continued to support the maneuver thesis include Frank van der Linden's *Lincoln: The Road to War* (1998) and William Marvel's *Mr. Lincoln Goes to War* (2006). Although somewhat dated, Richard N. Current's *Lincoln and the First Shot* (1963) has remained the standard and most thorough presentation of both sides of the argument.

Even more impressive evidence for the maneuver thesis than the opinions of the contemporary press, contemporary Southerners, contemporary biographers, and later historians has existed since 1861. Lincoln himself admitted to starting the war by maneuvering the South to fire the first shot. On May 1, 1861, Lincoln wrote a letter to Captain Fox in which he explained that Fox's failure to provision the fort was not worth fretting over. On the other hand,

according to Lincoln, Fox should take consolation in the fact that the failed mission was justified by the result. Also, on July 3, 1861, Lincoln boasted to Orville Hickman Browning, an old friend and political ally who recorded the conversation in his diary that his plan succeeded just as he had planned and the South's first shot gave the North a great advantage.

After the firing on Fort Sumter, Virginia, Arkansas, Tennessee, and North Carolina seceded from the Union. This brought the total of Confederate states to 11. Had Lincoln been more open to compromise, the war could possibly have been avoided. Also, had Lincoln not pursued his war policy, the South would not have fired on the fort. These last four states might have remained in the Union, and the South would have had far fewer resources to wage war. As it was, lines were being drawn. For example, Lincoln offered command of the Union Army to Robert E. Lee, but Lee declined the honor and went home to fight for the Southern cause instead. The costly war, which no one could have known would last four long years and result in so much death and destruction, pitted brother against brother. Years ago, historian Charles W. Ramsdell wondered if Lincoln's feelings of guilt for starting the war had not played a part in deepening his bouts of depression and encouraging his inexplicable charity toward his Southern foes.

References and Further Reading

Barbee, David Rankin. "Line of the Blood—Lincoln and the Coming of the War." *Tennessee Historical Quarterly*. 16, 1 (1957): 3–54.

Cothran, Helen, ed. *Abraham Lincoln*. San Diego, CA: Greenhaven, 2002.

Coulter, Ellis Merton. *The Confederate States of America, 1861–1865*. Baton Rouge: Louisiana State University Press, 1950.

Current, Richard N. *Lincoln and the First Shot*. Philadelphia: Lippincott, 1963.

Davis, Jefferson. *The Rise and Fall of the Confederate Government*. New York: D. Appleton and Co., 1881.

Marvel, William. *Mr. Lincoln Goes to War*. Boston: Houghton Mifflin, 2006.

Nicolay, John and John Hay. *Abraham Lincoln: A History*. New York: The Century Company, 1886–1890.

Potter, David M. *Lincoln and His Party in the Secession Crisis*. New Haven, CT: Yale University Press, 1942.

Ramsdell, Charles W. "Lincoln and Fort Sumter." *Journal of Southern History* 3, 3 (August 1937): 159–88.

Rutledge, Archibald. "Abraham Lincoln and Fort Sumter." *South Atlantic Quarterly*. (October 1935): 368–83.

Scrugham, Mary. *The Peaceable Americans of 1860–1861: A Study of Public Opinion*. New York: Columbia University Press, 1921.

Stephens, Alexander Hamilton. *A Constitutional View of the Late War between the States*. Philadelphia: National Publishing Co., 1868.

Swanberg, W. A. *First Blood: The Story of Fort Sumter*. New York: Scribner's, 1957.

Tilley, John Shipley. *Lincoln Takes Command*. Chapel Hill: University of North Carolina Press, 1941.

Van der Linden, Frank. *Lincoln: The Road to War*. Golden, CO: Fulcrum, 1998.

CON

What causes brothers, friends, relatives, sons, daughters, fathers, and mothers to rise in armed defiance against one another is a question that has plagued mankind since the conception of its civilization on the Fertile Crescent. In this war between neighbors, passion runs deep in the veins of its participants. Habitually the worst kind of war, civil war can call forth destruction of a society's ideals, peoples, and infrastructure, often leading to cataclysmic change. No matter the grandeur or strength of a civilization, a civil war will bring a country to its knees. With this in mind, the American Civil War brought the great democratic experiment to the point of an abyss. This war outright exterminated at least 618,000 of its citizens, desolated entire geographic regions, and destroyed a culture that had its roots even before the conception of the nation. Historians have debated how this war started ever since the final shots were fired. Initially, following the line that *victors write history*, the South was automatically blamed for its conception. However, with the passage of time, southern historians have vocally fought for a revision of this idea, often pointing to the focal point of the war, the then president Abraham Lincoln. They strive to demonstrate his instigation of the war. Although now a debate of only academic conjecture, this section will progress through the causes of the war and demonstrate that Lincoln, far from a perfect president, cannot be blamed for starting arguably the worst war in American history. It will conduct this exercise of histography by examining Lincoln's character and actions, the southern leaders and states, the economic, political, and social differences between the North and the South, and the human condition as a whole as it pertains to the causes of war.

Although it is impossible to know the true countenance of another's soul, it is plausible to re-create a picturesque idea by combining the impacts that nurtured their growth and the actions they took within the confines of the circumstances they endured. This technique of examination is no different, in that it scrutinizes the will behind one of America's most influential presidents.

Lincoln was born on a farm near Hodgenville, Kentucky, in 1809. Several years later his family moved to a 230-acre farm on Knob Creek and buried

Thomas Lincoln, Abraham Lincoln's brother, who died in infancy. By 1815, Lincoln was attending school with his sister Sarah, and a year later his family moved again, this time from Kentucky to Spencer County, Indiana. Soon afterward his mother died of milk sickness and his father remarried. From this it is possible to speculate that Lincoln, by age 10, had been raised on farms, witnessed two family deaths, and encountered continuous uprooting of his home life. As the sole male child of a farmer, he undoubtedly conducted household chores and watched his father tirelessly toil the land, which probably created a respect for hard work. The deaths in his immediate family brought him an early realization of mortality, and the constant uprooting and remarrying of his father created an awareness of inevitable change. All of these early childhood impacts would foreshadow motifs in his character as he grew into manhood.

By 1823 Lincoln's father joined an abolitionist Baptist church, and on August 2, 1826, his sister died in childbirth. Lincoln traveled to the South for the first time, helping his friend Allen Gentry with a cargo load to New Orleans. In 1831 he and his cousin and his stepbrother left home and built a flatboat, which Lincoln piloted to New Orleans. Also in this year, Lincoln participated for the first time in the electoral process by casting his first vote and began clerking in a store at New Salem. Now 22 years of age, the influences of the wrongs begot by slavery had been instilled in him by his father and undoubtedly by other church congregation members. Lincoln also showed his fascination with the South and his adventuresome spirit when he volunteered to pilot a flatboat to New Orleans. Additionally, even at this early age, he foreshadowed his desire to make a difference, tempered by the practicality of life, through his participation in voting while providing an income to provide for his basic needs.

In 1832 he became a candidate for legislature by championing river navigation improvements, changes in usury laws, and universal education, but he was defeated and placed 8th in a field of 13. In the same year, he was elected captain twice and enlisted and reenlisted seeking action in the Black Hawk War. He then became part owner of the store where he clerked by buying into it for $750 and became postmaster at New Salem. By 1834 he became a deputy surveyor and ran and was elected as an Illinois House Representative. However, a year later his sweetheart Ann Rutledge died. By this time Lincoln had begun to show his leadership qualities. His election to the rank of captain represented what others saw in him as being a capable leader who also had a willingness to be led, as demonstrated through his enlistment under other captains. Additionally, his election to the Illinois House showed he was gaining experience in the electoral process, even amid the ever presence of death that marked his life so tragically.

In 1836 Lincoln purchased from the government 47 acres of farm land and began his self-taught studies for practicing law. A year later he made his first attack on slavery by protesting an antiabolitionist resolution and became a law partner with John Stuart and Joshua Speed. In 1842 Lincoln married Mary Todd,

and a year later Lincoln's first child was born, soon followed by a second in 1846. All within this time frame, Lincoln symbolically connected himself to the land by purchasing acreage, found his occupational home in practicing law, and began his family life.

Also by 1846 he won the nomination and became a Whig congressmen and moved to Washington. Here he opposed the Mexican War, proposed an amendment abolishing slavery in Washington, was admitted to practice law before the Supreme Court, and was offered and declined governorship of Oregon. Four years later his second son died after 52 days of agony, but he also celebrated the birth of a third son. In 1851 his father died, and two years later his fourth son was born. In 1858 he gained national attention in debates with Douglas, and on November 6, 1860, he became the first republican to be elected president of the United States. Throughout this time, the idea of death and rebirth is again revisited to him as he gains national attention and eventually becomes the president under a newly organized republican nomination.

Amid reports of an assassination plot, Lincoln arrives in Washington in secret and is inaugurated the 16th president on March 4, 1861. On April 15, in a proclamation he called forth 75,000 state militia men to be activated and four days later initiated a blockade of the Deep South. He also suspended habeas corpus in several regions of the United States and increased the size of the regular army and navy. Touched by the death of his friend Colonel Ellsworth, on the fateful day July 4 he sent a formal war message to Congress. By this time Lincoln no longer commanded the initiative, and forces outside his control had now commandeered his circumstances, leaving only a trail of reaction and action by him and his opponents. This stage was set even before he took office, as the division between the North and the South had already brought the confrontation to a point of no return.

To dictate that Lincoln sought war or somehow pursued an agenda that he thought would lead to it is a highly ludicrous notion. This is because what primary resources exist all indicate that Lincoln grew up in a loving home and had a strong relationship with his stepmother and children. The numerous personal deaths he endured were probably the cause of his bouts of depression, which he suffered occasionally. Therefore, it is possible to conclude that he new firsthand the suffering that death brought and would, as he did during the Mexican War, oppose war as a means of politics. His opposition to slavery is well known, which dispels any chance of deception. His adventures on the flatboat incline many to believe that he had pleasant thoughts of the South. His study of the law and his numerous research and trial experiences made him an expert of the law. His apparent willingness to remain indifferent to church membership by allying himself to none dispels or at the very lest diminishes a notion of a dominant influence from religion in his politics. He was far from a radical in his belief that slavery was a moral wrong, as the majority of the population pursued the same course, as represented through his election. He never, in any statement or action, publicly or privately recorded,

worked to abolish slavery where it existed prior to the start of the war. Even before his tenure in office, the Confederate States of America had been created, forcing him to react without compromise or negotiation. As a Unionist he had no choice but to use force to bring back the wayward son into the fold as signified by the use of violence by firing first on Fort Sumter and seizing all federal properties in the South. In essence, he only called up the banner of arms in an attempt to maintain the peace and order. As a governor sometimes calls up the state militia upon its own citizens to restore order, so too did Lincoln use this necessary force. His only recourse would have been to allow the dissolution of the Union, which we could only speculate on as to the results, but would as a whole have inevitably negatively impacted the idea and the state of the *great democracy*. In Lincoln's words, which he quoted from the Bible, "a house divided cannot stand." Also, as stated in his Inaugural Address: "In your hands, my dissatisfied fellow countrymen, and not in mine, is the momentous issue of civil war. The government will not assail you. . . . You have no oath registered in Heaven to destroy the government, while I shall have the most solemn one to preserve, protect and defend it." Even after the final shots were fired, Lincoln worked to preserve the Union by opposing many of his political supporters in supporting a lenient reconciliation of the South to the Union. In essence, the research surrounding Lincoln's life supports the theory that he was an ambitious and hardworking man who endured losing many loved ones, suffered from bouts of depression, had a tremulous relationship with his wife, was a gifted orator and law practitioner, had a genteel spirit and firm convictions rooted in law and his own revelations of right and wrong, was compelled to react in preservation of the federal government, democracy, and the Union amid repeated personal tragedies, and in all instances as president prior to the Civil War lacked the initiative to create war, and thus could not have personally induced its start through his personal actions.

In contrast to this, southern leaders led their states onto a collision course that inevitably resulted in a divergent track that led away from the Union and the common good of the nation to a more personal venue of economic and social acceptance. Although slavery is commonly referred to as the cause of the Civil War by naive historians, more seasoned veterans of history state that is far from the case. However, most can agree that slavery was the lynch pin that held the antagonism between the North and the South, and without its existence, the war probably would not have occurred. The institution of slavery defined the South's peculiar economic and social condition, and the defense of it was paramount in keeping the status quo. As often happened throughout time, those in power pursue courses of actions that perpetuate their own place in society. In the 1860s in America, this was no different, proven by the simple fact that in the South, slave owners were the men who held the reins of power in the spheres of politics, economics, and social status.

How slavery became to dominate the South requires an explanation that is beyond the scope of this section; suffice to say, the South witnessed the rise of

small farmers to large plantations that needed an abundant and cheap labor force. With the termination of indentured servants, it became common practice to employ slaves to meet their needs. This combined with the economic rise of cotton on the world market, which required a labor-intensive growing and production process, and the South became dependent on the labor of slaves as a feasible economic stopgap to meet the global market demands. However, by 1860, because the importation of African slaves had long since been discontinued, the value of the slave had increased. This led to only approximately 46,000 Southerners owning 20 or more slaves, with fewer than 3,000 owning 100 or more, and only 12 owning 500 or more out of a total population of 5.5 million white Southerners. With this increasingly rare commodity, the prestige of owning slaves coincided with economic prosperity and social influence. In essence, becoming a large plantation owner became the southern American dream. As the future confederate president Jefferson Davis's father's story points out, starting as a small farmer, if one worked hard enough he could buy one or two slaves and increase his production. With that increased production he could buy more land and more slaves and eventually rise to prominence. It was no coincidence that the reins of power in the South were held by slave owners and thus their agendas became the agendas of the southern states. Under the guise of "states' rights" they sought to pursue the maintenance of the agricultural status quo of the South.

The most picturesque example of how the South pursued its agenda prior to the start of the Civil War can be found in one of its prominent leaders, John C. Calhoun. Calhoun could be called the South's political founding father from the 1820s to 1850s. The architect of the South's defense strategy, he developed a two-point defense system based on the protection of the minority, as in the political minority, of the southern states, based on the argument that the institution of slavery was as beneficial to the slave owners as it was to the slaves themselves. As an active political agent, Calhoun worked tirelessly in pursuing his agenda. For example, he wrote "South Carolina Exposition and Protest" in which he said that states had a constitutional right to nullify federal government actions they considered to be unconstitutional. He also championed that individual states had a legitimate right to leave the Union. Additionally, Calhoun defended slavery as a "greater good," by promoting the concept that the world existed on a racial social order that was run by self-interested humans. Blacks were inherently inferior, and the institution of slavery brought them a degree of civilization they were unable to attain for themselves. Furthermore, he asserted that the South's slavery system valued the black slave as an investment that was cared for, as compared to the North's wage-based system that enslaved people with wages that were barely sufficient to provide for their most basic of needs.

How southern leaders were able to incite the ultimate pinnacle of their separatist ideas over cooler political heads can best be explained by sympathizing with the average Southerner at the time. These usually hard-working people believed
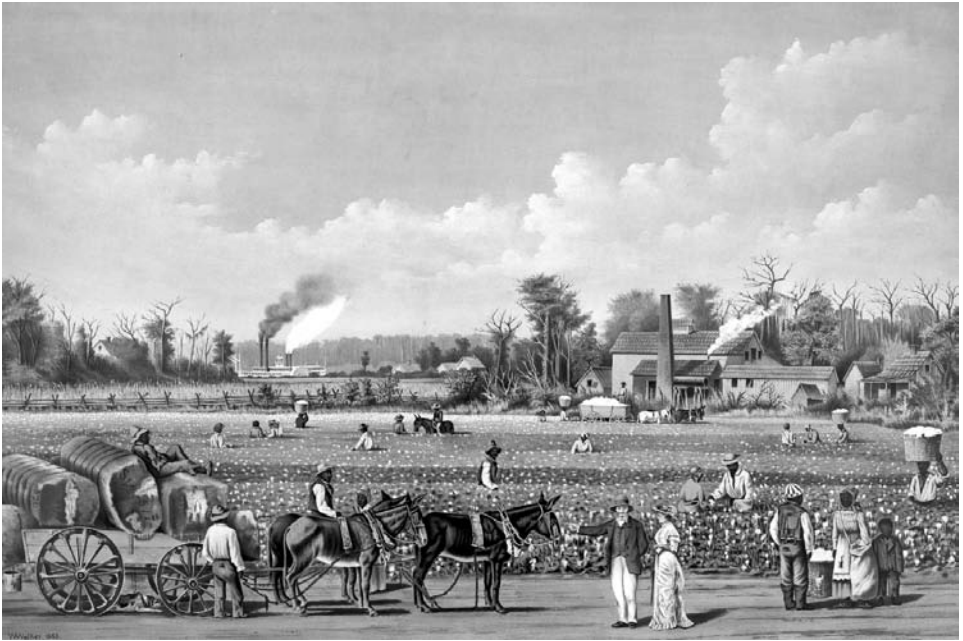
A 19th-century painting of a Southern cotton plantation near the Mississippi River. This romanticized depiction of plantation life shows the production of cotton through slave labor and the wealth of the Southern cotton aristocracy. (Library of Congress)

in the commonly held notion that blacks were inferior and the attainment of slaves was the path toward success, which was supported by biblical Christian beliefs. They complained because of the high tariffs imposed by the federal government, they traveled little and their knowledge of the outside world was that of sensational local papers and community gossip, and they felt besieged by what they believed was an increasingly overpowering northern-run federal government. The average Southerners were Jeffersonian agricultural followers who supported the separatist movement as a necessity for preserving their way of life. As a result they most often overwhelmingly supported the decisions by their legislators to dissolve the Union.

The differences facing the North and the South prior to the advent of the Civil War were paramount in lighting its powder keg. These differences can be divided into economical, political, and social gulfs. The disparity of economics between the North and the South was that of night and day. The South remained, for all intents and purposes, a rural agricultural state whose most important product was cotton. In fact, by 1850 the South grew more than 80 percent of the world's cotton, which represented 57 percent of the total U.S. exports in 1860. Additionally, crops such as tobacco, rice, and sugarcane were staples in many southern state economies. Although the South did have some industry and railroad, most notably in Richmond, Virginia, a center of slave trade and a railroad

hub, the South was far lacking in these areas as compared to the North. As a result, the South became an almost exclusive exporter of agricultural goods and importer of manufactured products. This naturally inclined the populous to favor low tariffs. In contrast, the North was fast becoming an urban population and the nation's center for manufacturing, commerce, and financing. Products such as lumber, textiles, industrial machinery, and leather were most important to the northern economy. Additionally, Northerners favored high tariffs, as this protected their goods from foreign competition.

The political differences between the North and the South polarized the nation into two competing branches that compromised and wrangled for dominance with each other. Initially starting with the conception of the nation, the North and the South from the very start began as two different societies. The North was formed out of the wilderness as an immigrant land for refugees seeking religious freedom; the mountainous terrain and many naturally protected harbors led to small tight-knit communities following the guidance of the "city on the hill" mentality, with large economic trading centers on the coastline. In contrast, the South originally started out with economic ventures whose many river systems allowed for ease of transportation and a more expansive population. As both were dominantly populated by English immigrants, they naturally formed political ties with each other, which axcumulated in the Union following the Revolutionary War. However, as the two grew apart, they fought each other for dominance, which resulted in historical conflicts and compromises such as the Missouri Compromise, Nullification Controversy, 1850 Compromise, Kansas-Nebraska Act, Fugitive Slave Law, and the Dred Scott decision. As a result of these events, the American federal political system fragmented into two camps: those who supported the northern view and those who supported the southern outlook. Eventually, it was the catalyst of American expansion that permanently divided the U.S. political landscape beyond repair. As the nation expanded westward following the Mexican War, determining what to do with the new states became the focal point for the new political battleground, with whomever winning dominance ensuring their view would become the course of the nation. It was not until the southern democrats foresaw final defeat in this venue, combined with the loss of the presidency to a sectional northern republican, that they envisioned the only way to preserve their political strength was to form their own self-government and thereby cement the differences in their relationships with the North.

The social differences between the North and the South were the culmination of their economic and political differences. The South was for the most part an open patriotically landed aristocracy based on capitalism. The gentry were supposed to act under certain constraints, and hospitality was the norm. A full third of the population was racially inferior with the vast majority of the rest of the two-thirds comprising small substance farmers. Extended travel was infrequent and slow, as little transportation infrastructure existed. In essence the South was

---

The Election of 1860

By the time the presidential campaign of 1860 began, the country was already thoroughly divided over the question of slavery. The election itself both reflected those divisions and, in its outcome, exacerbated them to the point of secession for many southern states. For their second attempt at the presidency, the new Republican Party nominated Abraham Lincoln of Illinois. By 1860, the country's divisions were reflected in the Democratic Party, with Northerners supporting Abraham Lincoln's longtime nemesis Senator Stephen A. Douglas and Southerners supporting a pro-slavery candidate, Vice President John C. Breckinridge of Kentucky. Remnants of the Whig and Know-Nothing parties formed a third party, the Constitutional Union Party, and nominated John Bell of Tennessee. On northern ballots, only Lincoln's and Douglas's name appeared, and on southern ballots, only Breckinridge's and Bell's name appeared. As a result, Lincoln, though he garnered less than 40 percent of the popular vote, won the presidency, as the North had over two-thirds of the nation's population. The divisions in the country and in the election itself proved too great for the Union to withstand, and the 'house divided against itself' could not stand.

---

reminiscent of Old World Europe. In contrast, the North was a bustling industrial center. Manufacturing and infrastructure development ran rampant. The citizenry was of an entrepreneurial spirit with progress fueled by ever increasing immigration. Additionally, as in the South, racism was also rampant, but it was directed toward the newly arriving immigrants instead of the blacks. From the altars of churches and within town squares, the abolition movement grew from a small radical group to mainstream thought. Although most believed in freedom of the slaves in the North, they did not necessarily believe in equality.

As a result of these differences, conflict and strife brought about by passions over slavery, economics, and politics inevitably occurred, as the two juggernauts of Americana fought for supremacy while moving in opposite directions.

Finally, it can be argued that wars in general are inevitable as a condition of human nature. The great human psychologist Sigmund Freud believed that humans have basic instincts of animal nature that dictate many of their actions. As a result of these basic instincts, wars are an inevitable result. This belief by Freud can be supported by the fact that archaeological evidence exists that points toward civilized violence starting to appear around 6500 BCE with the rise of urban centers. This is further supported by the logical conclusion that if people are prosperous, their less prosperous neighbors will envy and covet what the more prosperous ones have, which will eventually lead them to attempt to take either by force or coercion that wealth. This then forces a defense and increase of hostilities between the haves and have-nots, as resources are finite.

Another aspect of the human condition that leads to war is group narcissism, or the personal identification of an individual with a group that he or she idealizes.

This then results in a sense of belonging and an overinflated sense of importance as compared to others not in that group. In an evolved definition, this is nationalism or patriotism. Additionally, rage and projection are other causes of violence and war. Rage is anger about negative events that happen to an individual outside of his or her control. Projection, then, is focusing that anger onto a group or an individual, regardless of who is at fault, with reasons often fabricated or exaggerated in support of the belief. Furthermore, pseudo-innocence is another mentality that follows the human condition. It is allowing an individual to be lured into a false sense of innocence. The individuals and groups almost always believe that they are on the right side of the conflict. Their perspective is limited to their own, and their enemy's atrocities often become grossly exaggerated. Lastly, the mythical romantic lure of war as an adventure becomes a passage of manhood for many societies, luring recruits away from the hearth and home in pursuit of adventure, wealth, or defense of their home. Here the carnage that war brings is often purposely left out but rather portrayed as pompous parades that precede and follow victorious soldiers.

How does all this relate to the topic at hand? *Homo sapiens* have been at war with one another since the dawn of their civilization. This history of violence has not escaped even to the present day. The American citizens prior to the Civil War and especially in the South recognized their state governments, not the federal government, as the government of their homes. This group narcissism prospered peacefully in state allegiances. The presence of rage and projection was also paramount as the majority of the populations were underprivileged small farmers who toiled constantly on their tiny farms. Their lives obviously were a constant struggle, and they then projected their woes onto the federal government in protest of the high tariffs they were forced to pay for manufactured goods. And with limited travel opportunity and experiences, fueled by sensational local papers and gossip, they then projected their plight as being the result of the Northerners, who ignored laws that required them to return their runaway slaves and who, in their eyes, supported violent abolitionist actions, such as John Brown's Harper's Ferry raid. Also, pseudo-innocence was found in the psyche of the common Southerners as they fought for what they perceived as independence from a tyrannical federal government. They believed slavery was a better institution than the factory-driven economy of the North and that their way of life was a more authentically American ideal. In essence, they believed they were besieged by the northern-run federal government, and as such, had no choice but to withdraw from the Union and risk war as a possibility. Additionally, using selected passages from the Bible, they thought God was on their side, and since they were fighting against an aggressor on their own soil, they were in the right to defend their homes. Lastly, the short and relatively easily won Mexican War created a sense of euphoria that war was an adventure and that the upcoming war was going to be short lived and an opportunity for personal glory.

The American Civil War did not start through any one man's actions. There were many reasons why it started. In essence, it could be argued that it was inevitable given the plethora of differences between the North and South and their constant struggle for supremacy. Lincoln, as president prior to the Civil War, made no recorded statements on abolishing slavery where it already existed. Instead he believed, like many others at that time, that slavery was a dying institution and that it would eventually die on its own accord, as it gradually became less economically feasible. His only military action was in resupplying a few scattered posts that were still in federal hands. And in that he emphasized his only intent was to supply the necessities for life for the beleaguered garrisons. Even the former president James Buchanan agreed that the South had no right to secede from the Union, but because his term was nearly completed, he had little power to stop it. His attempt to call for 75,000 volunteers was a necessary risk, as many soldiers of the already pitifully small Union Army had converted to the Confederate cause. In essence, what is known about Lincoln indicates he would not have ended slavery in the South, probably would have worked to stop its spreading in the west, and condemned it morally, but he would have taken no action that he thought would lead to the dissolution of the Union. This is proven by the fact that he worked hard and risked alienating the border states in preservation of the Union by bringing the southern states back into the fold with the call of the volunteers. These volunteers were most likely Lincoln's last attempt to maintain the Union; however, his bluff was called by the Confederate states, and war erupted. Lincoln was backed into a corner even before he took office; the Confederate States of America had already been created, with all negotation attempts failing miserably and with the southern states only offering monetary compensation for federal property taken. In essence, the American Civil War started even before Lincoln took office. Southern leaders made the deciding vote and were determined to follow that course to the bitter end, rather than live under the Union banner. Facing this, Lincoln, a man who was witnessing his country's dissolution right before his eyes, could only react to his circumstances and not dictate them, as the nation was already torn apart by its own differences of economic, political, social, and psychological causes.

References and Further Reading

Carwardine, Richard J. *Lincoln: A Life of Purpose and Power*. New York: Knopf, 2006.

Donald, David. *Lincoln*. New York: Simon and Schuster, 1995.

Fehrenbacher, Don E. *Prelude to Greatness: Lincoln in the 1850s*. Stanford, CA: Stanford University Press, 1962.

Jaffa, Harry V. *Crisis of the House Divided: An Interpretation of the Issues in the Lincoln-Douglas Debates*. Chicago: University of Chicago Press, 1959; reprinted 1982.

Johannsen, Robert W. *Stephen A. Douglas*. Urbana, IL: University of Illinois Press, 1973; reprinted 1997.

LeShan, Lawrence L. *The Psychology of War: Comprehending Its Mystique and Its Madness*. Chicago: Noble Press, 1992.

Lincoln, Abraham. *Abraham Lincoln: Speeches and Writings, 1832–1858*. Edited by Don E. Fehrenbacher. New York: Literary Classics of the United States, 1989.

Lincoln, Abraham. *The Collected Works of Abraham Lincoln*. 9 vols. Edited by Roy P. Basler. New Brunswick, NJ: Rutgers University Press, 1953.

Neely, Mark E., Jr. *The Abraham Lincoln Encyclopedia*. New York: McGraw-Hill, 1982.

Weigley, Russell. *A Great Civil War: a Military and Political History, 1861–1865*. Bloomington: Indiana University Press, 2000.

Index

1900 C.E. TO PRESENT

POPULAR CONTROVERSIES

in

WORLD HISTORY

STEVEN L. DANVER, EDITOR

Popular Controversies in World History

INVESTIGATING HISTORY'S INTRIGUING QUESTIONS

Volume Four
The Twentieth Century to the Present

Steven L. Danver, Editor

ABC-CLIO

Santa Barbara, California • Denver, Colorado • Oxford, England

I

The Progressive movement in the United States and in other countries in the first decade of the 20th century represented a middle-class, conservative reaction against the rise of both big business and big labor that had created a status revolution. The middle class (journalists, attorneys, small town businessmen) supported reformers like Theodore Roosevelt and Woodrow Wilson in hopes of restoring what they believed were proper class relations.

PRO   Kevin Wilson
CON   Arthur K. Steinberg

PRO

The era between the turn of the 20th century and World War I was notable for a loosely connected progressive "movement," which attempted, variously, to expand the role of government, regulate the economy, and institute political and social reforms. From 1909 on, the time became known as the Progressive Era and has been generally understood as an era of enlightened, forward-looking, "modern" policies. This dominant interpretation of the era has meant that the word "progressive" has become synonymous with "liberal," antonymous with "conservative," and has come to denote "favoring or implementing social reform or new, liberal ideas." In contrast, "conservatism" has been defined as "cautious about change and innovation" and is often a synonym for reactionary. With the available vocabulary, arguing that the Progressive movement was in fact conservative appears to be not only unfashionable but actually a contradiction in terms.

In *The Republican Roosevelt*, historian John Morton Blum (1974) argues that Theodore Roosevelt was a conservative progressive, and in the process he offers a useful reconsideration of conservatism. Roosevelt believed the democratic process worked so long as moral individuals were at the helm, working through established institutions. Roosevelt valued established American institutions, which were forged in history, and he sought to preserve them. "But preservation," Blum explains, "depended on change." Change was thus not anathema to conservatism; in fact, change was necessary to preserve traditional treasured institutions and traditional values. Such gradual change included "change

I

obtained by adapting, managing, administering; change 'on suspicion.'" While the means to achieve necessary change took novel forms under Roosevelt, the goals of change were stability, order, and morality—quintessentially conservative ends. As Roosevelt once told a fellow progressive, "The only true conservative is the man who resolutely sets his face toward the future" (Blum 1974: 5).

By Roosevelt's own definition of conservative, it becomes clear that conservatism is not inherently opposed to change or reform, but a more precise definition is necessary. This section will present three aspects of conservatism and will demonstrate how the Progressive movement fit each of these definitions. First, the Progressive movement was conservative because it idealized the past and opposed radical social change. Progressive reformers inherited the American agrarian myth and worried about the consequences of the "closing" of the frontier. Like the populists, the Progressive movement attempted to restore the ideal social and economic conditions that they envisioned to have existed on a past agrarian frontier, but they rejected the populists' more radical propositions. Second, the Progressive movement was conservative because it sought to preserve existing social relations. The middle class formed the leadership of the Progressive movement and supported reformers like Theodore Roosevelt and Woodrow Wilson in hopes of restoring what they believed were proper class relations. Middle-class status anxiety was fueled by the rise of new elites, including corporate officials, industrial magnates, and political bosses. The middle class furthermore segregated themselves from those below them in order to preserve existing power relationships. Third, the Progressive movement was conservative because it aimed to uphold traditional values. Most progressives were raised in middle-class, evangelical Protestant households, and though many imbibed the secularism of their age, many retained a moralist tendency. Progressivism took the form of a moral crusade against perceived immorality, vice, and disreputable behavior, and progressive reformers sought to impose their middle-class morality on the upper and lower classes. Also, a politician like Theodore Roosevelt saw the promulgation of traditional republican moral virtue as necessary to reform the democratic system. Finally, the section will conclude with the argument that although the means of the Progressive movement were sometimes novel, the ends of the movement—as stated above—were essentially conservative. Novel means were necessary because society itself had changed. As historian Richard Hofstadter established in *The Age of Reform*, "Progressivism, at its heart, was an effort to realize familiar and traditional ideals under novel circumstances" (1955: 215).

As Hofstadter suggested, progressivism should be understood as a conservative reaction against the new social and economic conditions of late 19th-century America. The Industrial Revolution in the late 19th century transformed the United States from an agrarian into an urban nation. In 1860, farms composed 40 percent of the nation's wealth and in 1870 agricultural laborers still represented a majority of the nation's laborers. By 1900, farms accounted for

---

Richard J. Hofstadter, Historian of the Progressive Movement

Richard J. Hofstadter (1916–1970) was born at Buffalo, New York. His father, Emil, was a farrier who operated a retail store in Buffalo—his family was Jewish from Poland; and his mother, Catherine, was from a German Lutheran family. Orphaned by the death of his father when he was 10, he joined the Young Communist League while an undergraduate at the State University of New York and completed his doctoral thesis at Columbia University. He subsequently joined the Communist Party of the USA but left after a year following disillusionment caused by the Molotov-Ribbentrop Pact between the Soviet Union and Nazi Germany.

Influenced by Charles Beard and Beard's views on industrial capitalism, Hofstadter later became prominent for his own views, especially through his book *The American Political Tradition and the Men Who Made It* (1948), in which he first articulated his theory that Theodore Roosevelt was actually a conservative who was worried about "the people" becoming too powerful.

---

only 16 percent of the nation's wealth, and agriculturalists comprised only 38 percent of the nation's workforce. Farms continued to grow in number and in size as farmers moved westward, but the rapid growth of industry outpaced the agricultural sector such that by 1900 farm pay averaged $260 per year while nonfarm work averaged $622 annually. Between 1860 and 1910 the rural population doubled while the urban population increased by nearly sevenfold. Hofstadter aptly summarizes this trend: "The United States was born in the country and has moved to the city" (1955: 23).

The industrial growth of the so-called Gilded Age created a host of problems that were unprecedented in the American experience. Starting after the Civil War, the availability of industrial jobs drew millions of immigrants from Southern and Eastern Europe. These immigrants represented a change in type and in scale from previous waves of immigration. Immigration peaked in 1907 at over 1 million entering just that year, and by 1910, foreign-born persons accounted for about one-seventh of the population of the United States. While the labor pool grew, class conflict became more pronounced. Strikes proliferated as wages decreased in relation to inflation, and labor organizing increased. Along with the violence of the Homestead and Pullman strikes, the economic depression of 1893–1897, and the growing disparity of wealth, social upheaval seemed possible. The size of corporations was also without parallel. Just between 1898 and 1904, 234 trusts were incorporated with a total capitalization of $6 billion. The new firms included giants like United States Steel Corporation, Standard Oil, Consolidated Tobacco, and Amalgamated Copper. While capital became consolidated in fewer hands, the newly wealthy displayed their wealth ever more ostentatiously. According to one study, in 1892, 9 percent of American families owned 71 percent of all wealth (Hofstadter 1955: 136, 169).

Idealizing the Past

When Americans asked themselves why social and economic inequality and political corruption had arisen, they looked to the past. Historian David Wrobel argues in *The End of American Exceptionalism* that observers of American society, from the founders to the progressives, saw the existence of the frontier as the source of American exceptionalism. Wrobel explains, "the frontier of free or cheap land had always been the wellspring of American democracy" (1993: 51). The agrarian myth is the belief that uninhabited land and nearly limitless resources would allow most Americans to own land, and nearly universal land ownership was thought to make Americans socially equal and economically independent. Furthermore, as sovereign individuals, yeoman farmers were to be the backbone of a vigorous American democracy. In *Democracy in America*, Alexis de Tocqueville expressed his belief that "the soil of America is opposed to territorial aristocracy." In contrast to the European experience of landed aristocracy and landless peasants, a democratic division of land was believed to prevent the appearance of persistent inequality and class strife as in Europe. Thomas Jefferson expressed this belief in a letter to James Madison, when he wrote: "Our governments will remain virtuous . . . as long as they are chiefly agricultural; and this will be as long as there shall be vacant lands in any part of America. When they [the people] get piled upon one another in large cities, as in Europe, they will become corrupt as in Europe, and go to eating one another."

Jefferson's concern about urban corruption illuminates how the agrarian ideal was connected to democratic virtues. Hector St. John de Crevecoeur, a French expatriate, believed American political traditions were deeply imbued with values shaped by agrarian life. Upon reaching the New World, Crevecoeur, in his *Letters from an American Farmer* (1792) supposed the European "leaves behind him all his ancient prejudices and manners . . . [and becomes] a new man acting upon his principles" because of the wide availability of "unrented, untaxed lands." The frontier was widely believed to be a "safety valve" for the poor and discontented, and agrarian life was itself an education in democratic principles. The frontier made Americans socially equal, economically independent, and transformed them into virtuous democrats.

Progressive anxieties arising from the end of cheap agricultural land and growing economic inequality were foreshadowed by similar concerns that arose in the form of the populist movement. Like the progressives, the populists sought to restore social conditions that predated industrialism and commercial agriculture, a time when the agrarian myth was reality. Hofstadter writes:

> The Populists looked backward with longing to the lost agrarian Eden, to the republican America of the early years of the nineteenth century in which there were few millionaires and, as they saw it, no beggars, when the laborer had excellent prospects and the farmer had abundance, when statesmen

responded to the mood of the people and there was no such thing as the money power. (1955: 62)

The populists believed democratic reform necessitated corresponding economic reform, but the populists' plan to devalue the currency was merely a tool to restore the American economy to a state of free competition that prevailed before the Industrial Revolution. The populist movement rose and declined during the 1890s, but the reform impulse quickly spread to a larger constituency, one represented by various interests and with different goals, but nonetheless united by general principles that can be called "progressive." The populists and progressives shared anxieties about the consolidation of capital in a few hands, corruption of politics, and the commercialization of social relations. Also like the populists, the progressives sought reform, not to achieve utopia, but to turn back the clock.

Conservatism is a relative term. In late 19th-century America, progressivism was conservative in that, unlike populism, it did not seek radical social change. Progressive reforms were often backward-looking, half-a-loaf measures. Hofstadter writes:

Concerning the great corporations, the Progressives felt that they were a menace to society and that they were all too often manipulated by unscrupulous men; on the other hand, many Progressives were quite aware that the newer organization of industry and finance was a product of social evolution which had its beneficent side and that it was here to stay." (2002: 31)

Just as progressives objected to excessive centralization of wealth, they also took exception to centralization of political power in urban bosses, who were empowered by a constituency of new immigrants and whose corruption seemed to threaten the political culture of the United States. Progressives also saw unassimilated immigrants as a problem, but unlike populist nativism, progressives proposed less radical, more constructive treatment of immigrants. On the whole, middle-class progressives rejected radical socialism, but they also rejected the radical individualism that justified the creation of massive trusts and the neglect of public service. To remedy these problems, progressives tacked toward the middle, advocating morality in politics and what economist Richard T. Ely called "the golden mean" in economics. As Hofstadter argues:

Progressivism was a mild and judicious movement, whose goal was not a sharp change in the social structure, but rather the formation of a responsible elite, which was to take charge of the popular impulse toward change and direct it into moderate and, as they would have said, "constructive" channels. (1955: 163–64)

Like the populists, progressives inherited the agrarian ideal and also tended to romanticize agrarian life. The progressive journalist Jacob Riis said: "The

ideal, always in my mind, is that of a man with his feet upon the soil and his children growing up there. So, it seems to me, we should have responsible citizenship by the surest road" (Riis 1903: 24). The conspicuous centralization of wealth in trusts, the rise in labor unrest, and the growth of urban slums of Riis's time seemed to be symptoms of traditionally European problems. Many Americans attributed contemporary problems to the closing of the frontier, and some believed government would have to perform the function that had previously been served by the frontier. The progressive economist Richard T. Ely suggested in his 1903 book, *Studies in the Evolution of Industrial Society*, that the government would have to assume a new role in the economy to replace the frontier as the nation's economic equalizer. In a 1905 work, *Constructive Democracy*, William E. Smythe argued that the west could still serve as a safety valve, but only if the government provided funds to irrigate marginal, semiarid plains. Historian Frederick Jackson Turner likewise proposed the government could become the new guarantor of American democracy and struggled with the problem of assimilating new immigrants without the transformative frontier experience (McGerr 2005: 104).

One way that middle-class progressives actively addressed the problems of rural poverty in order to preserve the agrarian myth was the Country Life movement. Kenyon Butterfield, sociologist and president of the Massachusetts State College of Agriculture, was one of the movement's leaders. Butterfield said, "There is neighborliness in the country; there is intense democracy; there is a high sense of individual responsibility; there is initiative; but this over-development of the individual results in anemic social life" (quoted from McGerr 2005: 105). Progressives idealized agrarian life but believed unbridled individualism had to be tempered and a sense of community had to be fostered to best preserve the agrarian way of life. For instance, Ann Bassett, who grew up on a ranch in Colorado in the late 19th century, called for the federal government to regulate wealthier ranchers who were determined to drive the smaller ranchers off the land. While calling for greater governmental power seemingly violated her individualist ethic, Bassett argued that regulation was necessary just to preserve competitive capitalism and perpetuate yeoman farmers' economic opportunities. She recalled, "It was a privilege to live in a new free land where real democracy existed in a wholesome atmosphere, where people were accepted on their individual merits, and background or great wealth had small importance." There had been, she believed, "equal opportunity for development on all sides in an uncluttered America, before collectivism got a stranglehold on the nation" (quoted from McGerr 2005: 105, 147–48). But in her own time, large, unregulated cattle ranching operations used the individualist ethic to justify their refusal to pay the grazing tax or allow the government to count their cattle. Pitted against such foes, Bassett and other farmers decided that controlling big business did not represent a radical break with the past, but was in fact the best way to preserve the

agrarian way of life. While this logic legitimated antitrust legislation and government regulation, it did not succeed in returning America to the golden age of competition. Though government regulation of the economy was a novel means, securing the agrarian ideal was a conservative end.

Maintaining Social Order

Whereas the populists were often portrayed as radical anarchists seeking to overturn the social order, progressives were overwhelmingly middle class and respectable. Theodore Roosevelt's Bull Moose movement of 1912 was composed, according to William Allen White, "of little businessmen, professional men, well-to-do farmers, skilled artisans from the upper brackets of organized labor . . . the successful middle-class country-town citizens, the farmer whose barn was painted, the well-paid railroad engineer, and the country editor" (quoted in Hofstadter 1955: 132). Alfred D. Chandler, Jr., surveyed the Progressive Party leadership and found that they were largely urban, middle-class, native-born Protestants. Many were college-educated professionals, while others were businessmen; only one union leader, not a single farmer, and few white-collar managers were among the leadership. Chandler says:

Former president Theodore Roosevelt speaks to the Bull Moose convention in New Jersey in 1912. Dissatisfied with his successor, William Howard Taft, Roosevelt chose to run for president again, forming the Progressive (or "Bull Moose") Party when he did not receive the Republican nomination. Both Roosevelt and Taft lost to Woodrow Wilson in the election. (Library of Congress)

As lawyers, businessmen, and professional men, they worked for themselves and had done so for most of their lives. As individualists, unacquainted with institutional discipline or control, the Progressive leaders represented, in spite of their thoroughly urban backgrounds, the ideas of the older, more rural America. (Chandler 1958: 1465)

In another study, George Mowry found that the average California Progressive was "more often than not a Mason, and almost invariably a member of his town's chamber of commerce" (quoted in Hofstadter 1955: 145). Progressives were the petit bourgeois of their day, who, in the early 20th century, were experiencing a period of prosperity, but also—like the populist farmers—a relative decline in status.

The progressive leaders, Hofstadter argues, "were Progressives not because of economic deprivations but primarily because they were victims of an upheaval in status that took place in the closing decades of the nineteenth and early years of the twentieth century." Status and deference were diffuse until the mid-19th century. Local professionals, including lawyers, journalists, physicians, professors, merchants, and clergymen, commanded respect and wielded power in small towns until railroads created a national marketplace, corporations consolidated into massive trusts, and large cities became the new hubs of power and prestige. Hofstadter argues:

The old-family, college-educated class that had deep ancestral roots in local communities and often owned family business, that had traditions of political leadership, belonged to the patriotic societies and the best clubs, staffed the governing boards of philanthropic and cultural institutions, and led the movements for civic betterment, were being overshadowed and edged aside in the making of basic political and economic decisions. (Hofstadter 1955: 135–37)

Captains of industry, political bosses, and their allies surpassed the old, local elite in prestige as well as wealth. Middle-class professionals valued their traditional status, so the *nouveau riche* offended their sensibilities. Henry Demarest Lloyd wrote in *Wealth against Commonwealth*:

Our great money-makers have sprung in one generation into seats of power kings do not know. The forces and the wealth are new, and have been the opportunity of new men. Without restraints of culture, experience, the pride, or even the inherited caution of class or rank, these men, intoxicated, think . . . that they have created the business which has created them." (quoted from Hofstadter 2002: 35)

The old gentry viewed themselves as public servants; clergy, professors, and lawyers joined the Progressive movement as a reaction to the perceived immorality of the newly rich, the corruption of urban politics, the commercialization of society, the materialism and greed of the corporate world, and the vice of

urban slums. Progressive reforms represented a middle-class reaction against the new, monied elite. As Walter Lippman, a progressive journalist, observed, "The middle class . . . is the dominant power expressing itself through the Progressives, and through the Wilson administration. The middle class has put the 'Money Power' on the defensive. Big business is losing its control of the government" (quoted from Hofstadter 1955: 141).

If the Progressive movement sought to rein in the corruption of the new elites, progressives also distanced themselves from those below them on the social ladder. Segregation existed not only in the South, but it is a term than can be used to describe progressive attempts to draw boundary lines between the middle class and African Americans, Native Americans, immigrants, and the working class. Just as one goal of the Progressive movement entailed preserving economic opportunity, another was maintaining social stability. As historian Michael McGerr argues, "True to their mission to create a safe society for themselves and their children, the progressives turned to segregation as a way to halt dangerous social conflict that could not otherwise be stopped" (2003: 183). While residential segregation, economic discrimination, and legal disfranchisement of African Americans in the North and South was justified as protecting African Americans from abuse and brutality, racial segregation also benefited the middle class by maintaining unequal access to opportunity, wealth, and prestige. The same logic of delineation was at work in progressive policies that restricted Native Americans to reservations and disallowed them from becoming citizens. Moreover, limited immigration restrictions, interest in eugenics, and the growth of residential segregation by class and ethnicity all evince a middle-class desire to separate themselves and preserve existing social and power relationships.

The progressive middle class attempted to reverse its relative decline in status by remaking the other classes in its own mold. Michael McGerr describes the perspective of the progressive middle class: "The rich had their mansions and lavish entertainments, their divorces, yachts, and all the other trappings of consumerism and self-indulgence. Workers had their slums, saloons, festivals, their supposedly freer sexuality and expressiveness" (2003: 54). Middle-class progressives believed both the upper and working classes were morally depraved. The progressives' deep-seated notions of proper morality and public service justified their movement to reshape society. This missionary spirit was what the progressive Henry Demarest Lloyd alluded to when he wrote "The middle class is not to be exterminated, but is to absorb the other classes" (quoted from McGerr 2005: 74).

Upholding Traditional Values

The importance of evangelical Protestant religion and Victorian morality to middle-class progressives demonstrates the essential conservatism of the Progressive movement. Even though many progressives upheld the relatively "liberal" social

gospel, they remained within the bounds of the conservative Protestant tradition. Just because many progressives believed in the forward-looking social gospel does not mean that such progressives were not conservative, as the quote by Roosevelt suggests. Progressive moralists did not foresee a socialist utopia but rather sought a society that was in line with biblical principles. Even those progressives who were not devout Christians had absorbed evangelical Protestantism into their middle-class Victorian homes and their thought conformed to religious patterns. In fact, progressivism was deeply rooted in traditional moral values and often sprang from the American penchant for moral absolutism. As Richard Hofstadter argues, "It is hardly an accident that the generation that wanted to bring about direct popular rule, break up the political machines, and circumvent representative government was the same generation that imposed Prohibition on the country and proposed to make the world safe for democracy" (quoted from Crunden 1982: 40).

The traditional beliefs of evangelical Protestantism and more secular Victorian morality underlay the outlooks of various progressive reformers. For one, Frederic Howe was born into a middle-class family in a small Pennsylvania town to pious Methodist parents. Howe absorbed the morality of his pious Methodist parents, but was influenced by the growing secularization of late 19th-century American society. Still, the "morality of duty, of careful respectability" did not pass from his mind, as did the theological doctrines. Howe wrote:

> Early assumptions as to virtue and vice, goodness and evil remained in my mind long after I had tried to discard them. This is, I think, the most characteristic influence of my generation. It explains the nature of our reforms, the regulatory legislation in morals and economics, our belief in men rather than in institutions and our messages to other peoples. (quoted from Hofstadter 1955: 206)

In *Ministers of Reform*, historian Robert Crunden argues that the uniting belief of those who called themselves progressive was that they "shared moral values and agreed that America needed a spiritual reformation to fulfill God's plan for democracy in the New World" (1982: 40). Crunden illustrates how progressives, born between 1854 and 1874, learned the strict moral teachings of their parents, many of whom were republicans. Though raised in Christian, abolitionist households, their religion was the social gospel, which portrayed Jesus as more of a democrat than a messiah. Within this rigidly moral yet increasingly secular culture, in which the ministry was no longer an intellectually respectable profession, well-educated youth had no established career path. Progressives experimented with new professions, including social work, politics, and academia, because these professions would become respectable vehicles for their social consciences, as they attempted to reform corrupt individuals and institutions. Coming from similar traditions, Richard Ely, George Mead, John Dewey, and Woodrow

Wilson were pioneers in producing socially useful and morally satisfying work in academia. As a woman, Jane Addams had more difficulty finding such work.

As a girl, Jane Addams internalized her father's strict, Protestant moral imperatives, sense of social justice, and vigorous work ethic. When her father married an overbearing, frivolous woman, Crunden argues Jane's reaction against her stepmother made her into a progressive. "The foundations of Hull-House were laid in one woman's moral revulsion against privileged uselessness." From her father, Addams absorbed the moral skepticism of excessive materialism and frivolity. In reaction, Addams searched for a respectable outlet for her moral energies within the confines of Victorian vocational constraints on women. The result was Hull House, which represented the reification of Addams's social conscience. Crunden argues, "She was the first progressive to institutionalize her own psychological needs and thus the founding of Hull-House in 1889 marks an appropriate beginning for progressivism as an important cultural phenomenon" (Crunden 1982: 17, 19, 25). Just as internalized morality was a stimulus to individual action, appeals to traditional values worked on the collective level as well.

A central institution of the progressive era, muckraking journalism, reflected traditional values by voicing moral outrage at the persistent perceived immorality. Muckraking exposed political corruption, vice, and the abysmal living conditions in slums. Hofstadter argues the reality represented by muckraking "was a series of unspeakable plots, personal iniquities, moral failures, which, in their totality, had come to govern American society only because the citizen had relaxed his moral vigilance." Muckraking was moral exhortation aimed to enlighten the public conscience, and such appeals came naturally to the progressive moral outlook. Hofstadter explains, "The Progressive mind . . . was preeminently a Protestant mind; and even though much of its strength was in the cities, it inherited the moral traditions of rural evangelical Protestantism" (1955: 202, 204). Muckraking journalism gave evidence to the belief that politicians, corporate titans, and the wealthy had neglected their duty of public service and lost their sense of civic virtue.

With muckraking journalism, progressives pointed fingers of moral indignation not only at others, but also at the middle class itself. Protestant progressives were deeply troubled that social conditions deviated from those of their ideal agrarian myth, and they felt a degree of personal responsibility for the state of affairs. The deeply held ethos of personal responsibility, augmented by knowledge of lost virtue and neglected duties, produced a sense of guilt. In *The Shame of the Cities*, Lincoln Steffens wrote: "The misgovernment of the American people is misgovernment by the American people. . . . The boss is not a political, he is an American institution, the product of a free people that have not the spirit to be free." Steffens explained, "My purpose was . . . to see if the shameful facts, spread out in all their shame, would not burn through our civic shamelessness and set fire to American pride" (quoted in Hofstadter

1955: 208). Muckraking exposés like Steffens's *The Shame of the Cities* and Jacob Riis's *How the Other Half Lives* were intended as jeremiads, exposing wrongdoing to provide catharsis to the American character.

Progressives believed the surest method to reform the evils of society was to reshape individual human beings in accordance with their notion of proper morality. Attempts to reconstruct society in line with Victorian values included movements to ban liquor, eliminate prostitution, restrict gambling and pornography, and curb increasing divorce rates, all under the rubric of eradicating vice from public life. Carrie Nation was perhaps the most recognizable embodiment of the moral revulsion against vice. Her first husband had been an alcoholic, and her time serving as a jail evangelist with the Women's Christian Temperance Union taught her that vice was largely the result of intoxicating drinks. Motivated by her powerful faith and abhorrence of alcoholism, Nation sought to reform individuals by breaking up saloons and depriving society of sources of vice. Nation's activity inspired other prohibitionists to follow her lead and convinced authorities to enforce previously ignored prohibition laws in the state of Kansas. Middle-class religious fervor and assumptions about civic virtue shaped progressive reforms, but so too did the belief that persuasion alone was not enough. "Moral suasion!" Nation exclaimed. "If there's anything that's weak and worse than useless it's this moral suasion. I despise it" (quoted in McGerr 2005: 82–84). Some progressives, discontent with the results of exhortation and muckraking, resorted to compelling individuals to change their behavior, whether through force or through government action.

Along with settlement houses like Jane Addams's Hull House, evangelical Protestant churches were among the more important institutional promoters of progressive moral reform. Evangelical attempts to reach and reform the urban poor included Sunday schools, revivals, city missions, and settlements. The "institutional church" movement began in New York City in the 1890s and promoted a social vision of Christianity by providing education, recreation, and charity. By 1900, the growth of the movement gave rise to the Institutional Church League. The Salvation Army provided soup kitchens and secondhand stores, and the Young Men's Christian Association was another manifestation of the social gospel. But the evangelical nature of progressivism was perhaps most evident at the 1912 National Progressive Party Convention, where diverse progressive delegates attempted to forge a coherent platform. While this process was contentious, the overriding tone of the convention was that of a religious revival, laced with moral approbations and zealous hope for reform (McGerr 2003: 100–101).

Ends and Means of the Progressive Movement

The man who was nominated at the 1912 National Progressive Party Convention, Theodore Roosevelt, was in many ways representative of the conservative

nature of progressivism and the Progressive movement's tendency to use novel means to achieve conservative ends. Born into a wealthy family, Roosevelt developed an abiding belief that elites have a duty to perform public service and that government leaders have a responsibility to act morally. He was familiar from an early age with Christian ethics and their application in philanthropy and social work. John Morton Blum argues, "Roosevelt established for himself a stern moral code by which he rejected the amorality of business and public as well as of private life" (1974: 8). He shared the progressive bourgeois morality and encouraged other elites to live up to his moral standard. He was particularly critical of the new industrial and financial elites, whom he believed had abandoned traditional ethics. In the 1890s Roosevelt proclaimed, "There is not in the world a more ignoble character than the mere money-getting American, insensible to every duty, regardless of every principle, bent only on amassing a fortune, and putting his fortune only to the basest uses" (quoted in Blum 1974: 29, 33).

Though a conservative by nature and a republican by party affiliation, Roosevelt adopted the progressive mantle because he shared their conservative impulses. Like the progressives, Roosevelt argued in an article that while the "pioneer spirit" remained essential to the nation's continued prosperity, the closing of the frontier signaled that Americans needed to curb the individualist spirit in the interest of the "general welfare." Roosevelt also preferred segregation to violence as a resolution to, what he called, "the terrible problem offered by the presence of the negro on this continent" (quoted from Bucklin 1920: 166). Like many progressive reformers, in the realm of antitrust suits, Roosevelt attempted to redress the worst grievances but not reinvent or revolutionize institutions. In *The Triumph of Conservatism*, historian Gabriel Kolko explains, "Roosevelt was consciously using government regulation to save the capitalist system, perhaps even from itself, for the greatest friend of socialism was the unscrupulous businessman who did not recognize that moderate regulation could save him from a more drastic fate in the hands of the masses" (1973: 130). Like the Progressive movement as a whole, Roosevelt saw himself as an impartial arbiter mediating between capital and labor to reduce class conflict and to make the chances of competition more even. For Roosevelt, government regulation of the economy was, according to Kolko, "a means of preventing radical social change." Historian John Morton Blum agrees with this assessment of Roosevelt's ultimate goals. All Roosevelt's policies, Blum argues, "had one common, revealing objective: stability" (Blum 1974: 6). Roosevelt, like the middle-class progressives, idealized the past and sought to restore the social conditions of that ideal past, sought to maintain power relationships, and promoted traditional moral values—all in the name of stability.

If Theodore Roosevelt were the quintessential progressive politician, the classic statement of progressive ideas was *The Promise of American Life*, a book written in 1909 by Herbert Croly, a progressive political philosopher. The book

influenced Roosevelt's New Nationalism and prompted Felix Frankfurter to describe it as "the single most powerful single contribution to progressive thinking" (quoted in McClay 1999: 56). *The Promise of American Life* gave a heterogeneous reform movement a coherent set of ideas, but perhaps even more important, it also provided a national narrative that gave a larger context and justification for progressive reform. Croly described the American story as being driven by three impulses: a confidence in the virtues of agrarian individualism, commitment to a limited central government, and confidence in "the Promise of American life," or the gradual diffusion of democracy and steady amelioration of social and economic inequality. While a classical liberal ideology made these three tenets seem to reinforce one another, the changed circumstances of the late 19th century—the growth of massive corporations and increasing disparities of wealth—drew the ideology and the Promise into doubt. Croly believed that while too much emphasis had been placed on the first two principles, the conditions of the late 19th and early 20th centuries demonstrated that Americans needed to put relatively more emphasis on safeguarding the Promise.

The surest means to guarantee the Promise, Croly believed, was through a stronger central government and less emphasis on individualism. So while the means might appear unorthodox, the end was conservative. As Croly explained, Americans would have to use "Hamiltonian means," or a more powerful central government, to achieve "Jeffersonian ends," or democratic values and social and economic equity. The resulting expansion of federal regulation, empowerment of experts, and growth of bureaucracies may have represented a break with the past, but they represented a reaction, implemented to deal with the social, economic, and political consequences of the Industrial Revolution. Also, these new elites, whose job it would be to organize and rationalize the American economy, would ideally be restrained by an ideology of disinterestedness. Like Roosevelt's notion of public service, Croly believed these experts needed to put the public interest above more narrow interests. In the book's final page, Croly quotes Montesquieu's statement that "the principle of democracy is virtue." The circumstances of American life had changed, and the means of pursuing the Promise necessarily changed as well, though not entirely (McClay 1999: 64).

The confusion of ends and means is most notable in a classic revision of the Progressive movement. In *The Search for Order*, historian Robert Wiebe argued that the rise of the new elites favored by Croly was the defining feature of the progressive era. But this interpretation mistakes the means for the ends of the Progressive movement. While the expansion of government power was novel, it was one of many tactics in the Progressive movement's wide-ranging "search for order." The goals of the Progressive movement—re-creating agrarian conditions to revive economic equity, reestablishing conventional social relationships to maintain social stability, and restoring proper moral norms to promote a healthy political democracy—constituted enduring conservative ideals.

References and Further Reading

Bishop, Joseph Bucklin. *Theodore Roosevelt and His Time: Shown in His Own Letters*. New York: Charles Scribner's Sons, 1920.

Blum, John Morton. *The Republican Roosevelt*. New York: Atheneum, 1974.

Chandler, Alfred D. "The Origins of Progressive Leadership," *The Letters of Theodore Roosevelt, Vol. VIII*. Edited by Elting Morison. New York: University of Cambridge Press, 1958.

Crunden, Robert M. *Ministers of Reform: The Progressives' Achievement in American Civilization, 1889–1920*. New York: Basic Books, 1982.

de Tocqueville, Alexis. *Democracy in America*. London: Saunders and Otley, 1835.

Ely, Richard T. *Studies in the Evolution of Industrial Society*. New York: Macmillan, 1903.

Hofstadter, Richard. *The Age of Reform: From Bryan to F.D.R*. New York: Random House, 1955.

Hofstadter, Richard. "Do We Find the Roots of Progressivism in the Cities or on the Farms?" In *Who Were the Progressives?* Edited by Glenda Elizabeth Gilmore. New York: Bedford/St. Martins, 2002.

Kolko, Gabriel. *The Triumph of Conservatism: A Reinterpretation of American History, 1900–1916*. New York: Free Press, 1963.

McClay, Wilfred M. "Croly's Progressive America." *Public Interest* (Fall 1999): 56–72.

McGerr, Michael. *A Fierce Discontent: The Progressive Movement in America*. New York: Free Press, 2005.

Riis, Jacob A. *The Peril and Preservation of the Home: The William L. Bull Lectures*. Philadelphia: George W. Jacobs & Co., 1903.

Smythe, William E. *Constructive Democracy: The Economics of a Square Deal*. New York: Macmillan, 1905.

Wiebe, Robert H. *The Search for Order: 1877–1920*. New York: Hill and Wang, 1967.

Wrobel, David M. *The End of American Exceptionalism: Frontier Anxiety from the Old West to the New Deal*. Lawrence: University Press of Kansas, 1993.

CON

With the end of the Civil War and the Gilded Age, the United States entered the modern Industrial Revolution. This period centered on a collectivism that began in 1870 and closed with the American entry into the Great War, World War I,

when America turned to individualism. This collapse was partly the result of the success of the movements of various and dispersant groups. Various philosophies and political actions were taken in response to progressive efforts that molded a society that was more responsive to the needs of the majority, where little attention was given the wealthiest economic segments of a nascent industrial society. However, the atmosphere and climate-engendering progressivism were bleak for the average worker. The earmarks of the movement for this period were little governmental regulation, poor pay, abominable health conditions, and grossly inadequate economic facilities. The cause célèbre for the movement's success arose from despair, the intellectual belief that the worker could never attain the benefits being recently advertised in print throughout America during the Gilded Age. In juxtaposition to wealthy political and economic control, Americans of varied origins, backgrounds, education, and profession reacted to the societal inequities dominant throughout with a sense if social justice.

Elements of Liberal Progressivism

Conflict is present between all segments of society, and efforts taken to pursue changes in the relationships of these are difficult. An analysis of the attributes of liberal progressivism shows that its elements are similar to a liberal interpretation of fascism, even though the latter term is an anathema to many. Jonah Goldberg's (2007) *Liberal Fascism* provides an analysis of the subject beyond the scope of this section, but it is worth reading for a comparison of their respective elements. The tome argues that fascist elements in government existed well into the administration of Franklin D. Roosevelt. However, Goldberg posits that many do not understand the terms used.

The Progressive movement had tentacles that radiated into every sphere of American society. As a response to Victorian society and the Gilded Age, anger and hopelessness provided the hallmark through which middle-class Americans, educated and idealistic, launched their campaigns to change society. Help would be offered to both coal miner and slum tenant. Municipal regulation, moral indignation, and the value of human life, not merely existence, fell within the purview of the idealistic liberal Americans who chose to wage a war of progress devoid of the shackles of political, economic, social corruption, and stagnation of upward mobility caused by the robber barons and their minions.

These latter, the wealthy, believed in individualism and that they had earned their positions because of their exceptional ability, talents, and even origins, including religions. People such as Andrew Carnegie believed that an unfettered government was responsible for America's progress and that regulation of any kind would result in irreversible damage to society. They did not appreciate the cultural and political tensions created by their conduct. Progressivism sought to remove this class tension through the use of the intellect. The hope was that

---

"Fighting Bob" La Follette and Progressivism

The politician most identified with the Progressive movement was Robert Marion La Follette, nicknamed "Fighting Bob." The governor of Wisconsin, and later republican senator from Wisconsin, he contested the 1924 U.S. presidential election and managed to win not only Wisconsin, but also 17 percent of the national vote, standing as a candidate for the Progressive Party. He certainly would have achieved a higher result had it not been for a split and the failure of the progressives to nominate a single candidate.

From a French family who had lived in Wisconsin for four generations, Robert La Follette championed "the little man" against the large companies, especially the railroads, the banks, and the electricity companies. He also opposed U.S. involvement in World War I, and supported U.S. noninvolvement in the League of Nations. As a member of the U.S. Senate for 19 years, some historians regard him as one of the most influential senators in U.S. history. He died on June 20, 1925, seven months after his failed presidential bid.

---

educated people would, and could, rise above their petty differences and concerns for the good of society.

Conversely, the poor, lacking in power, economic or political, resorted to mutualism and cooperation for their collective defense. Families remained united, with their members working for the benefit of all. Their living conditions lacked most conveniences as is illustrated in Jacob Riis's (1890) *How the Other Half Lived* or Jane Addams's work in Hull House, assisting and protecting the needy of society. The proto-proletariat worked long hours in unhealthy conditions for low wages. Education, so necessary for upward mobility, was denied such a person because of the need to contribute to the family's well-being, a situation that only further magnified the differences between classes. The resulting resentment fueled the move toward collective action. Only the agrarian sector managed the shibboleth of independence and control over individual destiny. Besides the agrarian angst, the declassed American worker felt economically defenseless. Upton Sinclair's (1906) *The Jungle* illustrates the problem; labor was not yet organized, although nascent organizations became increasingly popular; John Wayne's rugged individualism had little currency for the majority of labor, whether skilled or unskilled.

Equal to this problem was the diversity caused by immigration. Religion and national origin made the problem of organization worse. Several ideologies rose to justify the slanted lack of economic and political control over society. The robber barons offered theories, such as the gospel of wealth, to affirm their places over the average workers. The gospel told everyone that God had ordained that work—a Protestant belief—had determined one's place in society and that God had prescribed a person's place and role in society.

Poor residents of a New York City tenement eke out a living making neckties in a local workshop during the late 1880s. Photographed by Jacob Riis, such makeshift sweatshops were common during the late 19th century. (Library of Congress)

The God-fearing and quasi-educated members of the middle class evidenced their lack of security and dissatisfaction as America moved from the Victorian to the Progressive Age. Members of the middle class were psychologically well adapted to the new theories of cooperation. Families learned to work together domestically, and because they were educated, middle-class members looked beyond individualism to cooperative approaches to their social malaise. Instability led to prostitution and unstable marriages as well as increased alcohol consumption, indicative of a decline in social mores, and the progressives viewed this as an opportunity to disseminate their movement goals for America.

Jane Addams's Hull House pursued radical change in society and epitomized the new age. Domesticity suffered; women were dissatisfied with their traditional roles and now sought something more fulfilling than their parents' roles. Many women, feeling the emptiness of their lives, joined the movement for more and better. They pursued diverse avenues and opposed the restrictions of the Victorian Age. The broadening of opportunities offered post-Victorian women new horizons, many of which their parents found unattainable: education and the power to refuse economic independence through marriage.

The freedom given middle-class women allowed them to evaluate their male counterparts, as female leaders of the Progressive movement acquired

increasing political power. For many, the result was not positive, and divorce and the subsequent independence it gave women became the harbinger of single-parent families. The Industrial Age contributed to the malaise and role conflicts that marked the middle-class family. Job uncertainty and the need to provide more produced emotional and familial disorder, creating a social climate that affected the Progressive movement's development.

Welfare Capitalism

The aforementioned domestic disasters, reflective of the decaying general economic and social systems, were justified through the use of welfare capitalism. Because of Victorian individualism—as Americans increasingly were economically reduced in power—progressive collective efforts gained more appeal. The character Julian West in Edward Bellamy's *Looking Backward: 2000-1887* (1888), a popular novel, described the conditions of most Americans in the protoproletariat segment of society. These conditions, in conjunction with the literature of many educated middle-class writers with a sense of social conscience—the muckrakers, allowed average Americans to explore the conditions under which they lived and resist justifications for and expressions of social Darwinism.

These terms were used to justify individualism rather than cooperative efforts to establish equilibrium among the all-economic elements in society. The rash of events such as the Haymarket Affair, the bombing that took place at a rally in support of the strike against the McCormick Harvesting Machine Company at Chicago's Haymarket Square in 1886, and the Pullman strike, the labor dispute between unions and railroads in 1894, publicly and increasingly threatened individual Victorian equality and led to cooperative action through labor unions, inviting a movement the declassed American worker found beneficial. Religious organizations supported many of those advocating change, even by violence. They indicated that the Gilded and Victorian ages were passé. However, the middle class's commitment to religion lessened. People discovered that the freedoms acquired through the Progressive movement allowed for more leisure time; the Protestant work ethic had less relevance to Americans than it had to their ancestors.

Seeking a Political Platform

With the inchoate successes of public awareness, progressives now sought a political platform for their various planks. Socialism, despite being accused of embracing anti-God and anti-American views, provided a potential way for defining the umbrella of a political movement committed to the middle- and working-class Americans. However, for various historical and ethical reasons, this "ism" was not accepted, and soon events such as the Palmer raids, the anticommunist

raids and deportations instituted by U.S. Attorney General A. Mitchell Palmer during 1918–1921, and the deportation of socialists alleged as enemy aliens strengthened the opposition to it. Government regulation in every segment of American life called for supervision of all sectors of the economy. Specifically, muckraking novels like Sinclair's *The Jungle* or Riis's *How the Other Half Lived* led to an awareness of the actual situation that prior rationalizations, secular and religious, and justifications had little currency to the actual situation that existed. The climactical opposition to any term identified as socialism led to the creation of a new identification not associated with the Bolsheviki or Karl Marx.

Faced with the difficulties in pursuing a comprehensive progressive plan for society, the necessity for governmental intervention, on all levels, became tantamount. They recognized that government from the municipal corporation to federal agency must join the battle to make America fit for all her people. This realization amounted to a revolution by the progressive middle class. They dedicated their efforts to every aspect of American life without any exclusion of class, occupation, sex, or race. This liberalism became the hallmark of a movement that would change America.

Control of Big Business

Control of big business was an essential arm of the progressive crusade. Given the consolidation of businesses and the lack of middle-class power, the progressives sought federal assistance in the battle for regulation. They aimed at national legislation but ran into the conflict of traditional federalism and state efforts to control national business enterprises. Businessmen were able to take advantage of the interstate and intrastate jurisdictions. Any confusion caused by this delineation aided big business and prevented progressive efforts to develop regulations for such sectors of the economy as railroads, food, and insurance. Hopes of building new associations or unions existed especially because progressives believed in environmental factors: bad food, unsanitary conditions, lack of education, and so forth. Presumably, such changes would reduce criminal conduct. These hindrances, it was thought, would retard the creation of a new man better able to function effectively with society.

Led by Carry A. Nation, a feminist, efforts to build a new nation were supplanted by her drive to construct a brave new man. She was not an armchair progressive but was physically committed to the cause. As a religious conservative, she offered a litany of sins, the elimination of which would be redemptive: all forms of gambling, pornography, contraception, divorce (as today, the male was considered the cause for the separation), and breaking the sabbath, and so forth. She reserved her greatest attack for alcoholism, which caused so much depravity. Progressives sought to create the new man free of sin, a family man. Nation used violence as a tool wherever and whenever she deemed it necessary.

The middle class bore these economic and social burdens because they had more education, idealism, and disposable income. Urban workers and wealthy citizens showed little interest in being morally molded; they tended to be more independent and willing to oppose actions that affected their private lives. Members of the middle class of society became muckrakers and advocates for progress in seeking the betterment of society. With these tools, concerned members of the middle class viewed conditions with disgust and strove to transform America, whether it liked it or not. Some have described this as the radical center of reform, but how does one define radical?

The hope of reshaping America and creating a more equitable society was the core issue of the progressives. The more politically observant recognized the relationship between business control of the federal and state legislative processes and the alliance with big business. Legislative reforms were viewed as the avenue through which the progressives could mold the new man or woman. Environmentalists, social scientists, and naturalists appreciated the connection between the well-being of the individual and a more equitable society; numerous muckraker authors illustrated the effects of society on our genetic composition. Josiah Strong, a preacher of the social gospel, lectured about the implications of a bad and unhealthy environment and the human results. Religion and evolution, justifications adopted by the upper strata of society as justifications for the conduct of man or woman were relegated to the trash can.

Jacob Riis's (1903) *The Peril and the Preservation of the Home* offered a more chilling account of economic conditions and their results. At the same time, evangelical Protestants and other religious groups made efforts to legislate morality; The Anti-Saloon League, the white slave trade, and Society for the Prevention of Crime were but a few of the organizations attempting to gloss a dysfunctional society with the moral glaze illustrated by the demise of the prohibition amendment to the federal constitution. Morality and the inability to govern their own lives made concern for the lower class an important issue. The reformers were neither fascists nor socialists within the common definitions of these words, however; they did not seek control but a more equitable opportunity for all Americans through unionization or organization, even though at times their methods might have been suspect. The organizational emphasis stressed by the progressives, while attractive to urban poor and middle class, was not as attractive to the farmers and other groups who cherished individualism.

Education as Concern

An area of general concern was education, which became a cause célèbre to the progressives. They transferred much of their general concern about society to the specifics molding and affecting our culture. Dysfunctional families, poor housing, low wages, and insufficient food fell within the purview of the progressives as

factors that had to be resolved in order to have efficient and progressive educa-
tion. In apparent response, the U.S. Supreme Court ruling in *Hammer v. Dagenhart*
defeated their attack against child labor. The Court ruled that child labor laws in
interstate commerce were unconstitutional. And, since the nation was increasingly
becoming one economic unit, portends for the future were glum for progressive
attempts to educate and uplift society.

Yes, several courts supported legislation that aided the progressives, such as
in juvenile courts, but state courts provided more assistance than the federal sys-
tem. A movement to protect children was part of progressivism. However, as
with immigrant newcomers, the progressives sought to substitute "American"
society and its values for the customs of the various European states, especially
the southern and eastern European states. Social workers concentrated on school-
ing, education, and hygiene unmindful of the needs of the immigrants. With the
latter, social Darwinism was injected into the climate by those opposed to pro-
gressivism. The Darwinian proponents vied with the social gospel elements in
society, and the motives of the liberal reformers came into question.

Justification for laissez-faire conduct by business with its concomitant opposi-
tion to liberal Progressive movements strained efforts to establish collective activ-
ity, whether unionism or association, among labor. Many of those imperiled by the
system came from European countries referenced above. Ethnic conflicts and the
natural distrust of groups not familiar with one another made progressive efforts
more difficult. The tenuous relationship between labor and capitalists hindered pro-
gressivism. Progressivism was hindered by the American Federation of Labor's
(AFL's) insistence on a craft union to the exclusion of those less fortunate. Wel-
fare capitalism became the vogue as a means of eradicating public efforts to
attempt collective action; an increasing number of workers supported these efforts
and disavowed collectivism. A less congenial approach for the workingman was
led by Frederick Taylor, a scientific management expert, who provided a sense of
individualism for labor; one's earnings were reflected by his or her production.

Influencing or conducting policies that aided labor and society became a
problem because some unionists were elitists. The AFL attempted to influence
many of the agencies that the progressives targeted: municipal government, state
labor federations, and unions on all levels. When combined with the concept of
welfare capitalism, only tragedies such as the Triangle Shirtwaist Factory fire of
1911, which killed 146 mostly female and mostly immigrant workers, and the
support of immigrants from Eastern Europe allowed the Progressive movement
to move forward.

Progressives achieved success when they supported provisions aiding
women and children to function in the American industrial society. Discourage-
ment arising from Supreme Court decisions that hindered the progress foreseen
by the middle-class reformer only proved to be an obstacle. Low wages, poor
working conditions, and the inability to obtain an education because children

had to work to support the family saw changes in the state level that provided encouragement. *Muller v. Oregon* recognized the inequities of the time and mandated improved working conditions for women as the courts expanded their roles in sanctioning better working conditions. Unfortunately, the emphasis on women's working conditions led to a return of Victorian expressions of the role of the woman. Conversely, American labor sadly suffered under the tutelage of the Fuller Supreme Court. The justice and the majority of the associate justices on the Court spoke of contractual issues, property rights, and federalism.

Ideological Problems

Progressive philosophy faced ideological problems; the various interest groups desiring change could not form a united front, socially or politically. While some progressives wished to regulate business for the common good, others wished to dismember the large corporations exemplified by the robber barons. Middle-class advocates, whether teachers or secretaries, each viewed the social malaise differently. New arenas for conflict developed in the growing American industrial society. Now railroads, technology, governmental policies, and social Darwinism were all in conflict. Theodore Roosevelt, a supporter of change and the presidential nominee of the Progressive Party, was unable to constrain the movement's activities, partly because many feared a loss of individual liberty because the law increasingly regulated society. Meat regulation, drug control, and other forms of legislation were instituted, but the mythical fear of the loss of individuality caused problems for the progressives. These latter opposed progressive segments, hoping that governmental regulation would expand. William Jennings Bryan, the Scopes trial attorney and progressive presidential candidate, was quoted in the *Congressional Record* in 1908 as saying that "public ownership is necessary where competition is impossible."

Conclusion

Progressive Americans viewed their efforts to "correct" as liberation from the politics, economics, and individual lifestyles of the past and an ushering in of a utopia like Edward Bellamy's (1888) *Looking Backward*. Americans looked forward to progress, better living conditions, and a better life. These hopefuls opposed the supporters of big business and the religious conservatives who contested science and the moral code befitting a modern growing society. The progressives, through their liberal policies, encompassed all aspects of growing industrial America. One of the more pronounced changes was the status of women who removed the religious, economic, and legal fetters imposed over the millennium. Conditions for black Americans were not well addressed because they lived in the South prior to the Great Migration; oriental Americans suffered the same because the members

of these groups didn't live in large industrial urban centers. It would take a need for workers in all fields of the military budget, World War I or World War II, before large numbers of these minorities were under the looking glass. Americans now questioned terms such as social gospel, social Darwinism, gentleman's agreement, gospel of wealth, welfare socialism, socialism, and so forth.

Sadly, with the commencement of World War I, President Woodrow Wilson had to create a wartime economy as a result of many progressive programs. Efforts suffered, only to be reborn under the administration of Franklin Delano Roosevelt.

References and Further Reading

Bellamy, Edward. *Looking Backward: 2000–1888*. Boston: Ticknor, 1888.

Goldberg, Jonah. *Liberal Fascism: The Secret History of the American Left, from Mussolini to the Politics of Meaning*. New York: Doubleday, 2007.

Riis, Jacob. *How the Other Half Lives: Studies among the Tenements of New York*. New York: Scribner's, 1890.

Riis, Jacob. *Peril and the Preservation of the Home*. Philadelphia: Jacobs, 1903.

Sinclair, Upton. *The Jungle*. New York: Doubleday, 1906.

2

The captain of the ship *Californian* was guilty of gross negligence in not coming to the rescue of the survivors of the *Titanic*.

PRO   Tim J. Watts
CON   Elizabeth D. Schafer

PRO

In April 1912, the ocean liner RMS *Titanic* made its maiden voyage from Great Britain toward New York. On the night of April 14, she struck an iceberg. After only two and a half hours, *Titanic* sank, taking over 1,500 passengers and crew with her.

The disaster was made more tragic by the failure of one nearby ship to render assistance. The *Californian* was within 10 miles of *Titanic* when she struck the iceberg. Her wireless operator was in contact with *Titanic* before the collision. Officers and men on board the *Californian* were able to see the *Titanic* and saw distress rockets fired off to attract help. Still, *Californian* failed to make any attempt to assistant *Titanic* or to rescue survivors until the next morning. Her captain actively took steps to cover up his ship's presence near the *Titanic*. Only after the inquiries conducted by the British and U.S. governments confirmed the presence of *Californian* on that fateful night did her captain grudgingly admit the ship might have been in a position to assist.

The Failure of the *Californian* to Come to the Rescue of the *Titanic*'s Survivors

By tradition and international law, sea captains are required to assist other ships and sailors in trouble. On the night of April 14, 1912, Captain Stanley Lord failed to render aid to *Titanic* and its passengers and crew. The evidence indicates that he received reasonable indications that another ship was requesting help. Although Lord later claimed that he could not be certain of such a request, he unquestionably failed to take any action to clarify the situation and later tried to deny any responsibility.

Lord's own officers later indicated they felt he did not take their observations seriously. Both the British and American inquiries assigned him some responsibility for the great loss of life as well. By failing to sail toward *Titanic* until it

25

Stanley Lord, the Man Who Saw the Iceberg Coming

Stanley Lord was born on September 13, 1877, at Bolton, Lancashire, in the north of England. His father, James Lord, was a commercial clerk and later a cashier at a cotton mill. Living at 9 Hampden Street, Little Bolton, Stanley grew up with four older brothers, and indeed he himself started work as a clerk at the age of 13. However very soon afterward, he went to sea as a cadet and later gained his second mate's certificate and then his master's certificate—the latter at the age of 23. In the following year he was living at Walton on the Hill, Lancashire.

In 1897 he had started working for the West India and Pacific Steam Navigation Company, and after they were taken over by the Leyland Line, with that company. On April 5, 1912, he spotted the iceberg and instructed his wireless operator to send a message to the *Titanic*.

Lord continued serving in the British merchant navy throughout World War I and retired in March 1927. After the release of the film *A Night to Remember* (1958), he was angered by the mention in that film that the *Californian* was closer to the *Titanic* than was actually the case. A legal battle resulted. Captain Stanley Lord died on January 24, 1962, at Wallasey, Cheshire, aged 84. He was buried in the New Brighton Cemetery in Merseyside.

---

was too late, Lord indirectly contributed to the many deaths. He was removed from his position and lived under a pall of disfavor for the rest of his life.

The Facts

Stanley Lord had been a ship's captain since 1906. He was an experienced seaman, even though he was only 35 years old. Lord had been at sea nearly continuously since enrolling as an apprentice with J. B. Walmsley & Company in 1891. During that time, he had proven himself as a conscientious and competent officer. Like many ships' captains, Lord could apparently be a domineering and strong officer who did not encourage input or uninvited suggestions from his officers or men.

By 1912, Lord was employed by the Leyland Line, which operated 30 ships. In March 1911, he took over his largest command, the 6,223 ton *Californian* (sometimes written as the *California*). The *Californian* normally carried up to 47 passengers, along with its main load of freight. She was not a passenger liner and was relatively slow. Although her best speed was approximately 13 knots, *Californian* ordinarily sailed at a slower, more economical pace. During Lord's first year in command, the *Californian* traveled the sea routes between North America and Great Britain. On April 5, 1912, *Californian* set sail from London, bound for Boston. She carried no passengers, but had a full load of freight. Superstitious sailors may have been concerned over a traditionally unlucky Friday sailing.

Californian's course toward Boston took her along the major sea routes across the North Atlantic. She was equipped with a new radio set, known at the time as a "wireless." Although this was only *Californian*'s second voyage since having a wireless fitted, Lord had been pleased with the results. The radio had brought messages about sailing conditions from other ships. In April, one of the most important warnings was in regard to ice and icebergs. Sailors such as Lord were very conscious of the dangers ice presented to their ships and avoided it as much as possible. For example, *Californian* had received a warning of three icebergs from another ship bound for Boston earlier on April 14. She altered course and passed within five miles of the icebergs around 6:30 P.M.

Concerned about the dangers ahead, Lord ordered extra lookouts and planned to spend the night on the bridge. He assumed control of the ship around 8:00 P.M. At 10:15 P.M., Lord detected an unusual brightness on the horizon. Experience convinced him that this was a condition known as "ice blink" among sailors. Field ice and ice floes on the water reflected the light from stars, making the reflection more visible at a greater distance than the normal reflection off water would have permitted. Lord ordered the ship stopped immediately. *Californian* slowed to a stop with its bow pointing to the northeast. After examining the conditions around him, Lord decided it would be safest to remain stopped for the night. In the morning, he could see how thick the ice was and whether it was possible to find a passage through, without causing damage to the ship.

Lord later reported the night as being extremely clear and the weather calm. The visibility was very good. He prepared to return to his cabin since the ship had stopped and all seemed secure. At about 10:30 P.M., Lord noticed a light to the southwest, which he believed to be the masthead light of another steamship approaching from the east. He called it to the attention of Third Officer Charles V. Groves. Groves stated that he believed it was a star, which seemed to satisfy Lord.

Before retiring to his cabin, the captain walked around the ship to be certain everything was secure for the night. Lord spoke with Cyril F. Evans, the *Californian*'s wireless operator. Evans reported that he had been in contact with the *Titanic*'s operator earlier that day. *Titanic* had confirmed it received a message from the *Californian* about the three icebergs seen earlier. Lord ordered Evans to tell *Titanic* that *Californian* was stopped for the night because of the ice field. When Evans tried to do so, the *Titanic*'s operator cut him off. He told Evans that Evans's signal was interfering with messages being sent from *Titanic* to New York. These were known as "Cape Race" messages, because they were sent to a radio relay station on Cape Race. The relay station forwarded the messages from passengers to individuals on land to let them know of their impending arrival. Evans could judge that *Titanic* was relatively close from the strength of its radio signal. He turned in for the night around 11:30 P.M. No one else was available to monitor the radio messages.

Before Lord went to his cabin at 11:45, he noticed a ship to the south of *Californian*. At the time, he apparently believed it was a steamer about the size of the *Californian*. Other members of the *Californian*'s bridge crew also saw the ship. They all agreed that it was stopped, much like the *Californian*. Lord told Herbert Stone, who assumed control over the *Californian* at midnight, to watch the other ship and to notify Lord if it seemed to be drifting toward the *Californian*.

The other ship was the *Titanic*. It struck an iceberg at 11:40 P.M. and slowed to a halt. At the time, it was probably about 9 or 10 miles away from *Californian*. The *Titanic*'s bow was pointing nearly directly toward the smaller ship. Eyewitnesses on *Titanic*, especially Fourth Officer Joseph G. Boxhall, noticed the lights of a ship to the north. Boxhall later testified that he was certain the ship was not a small sailing ship, as was sometimes claimed, but a steamer. He and Captain Edward J. Smith were the only officers to see the mystery ship through binoculars. Smith told the crew manning the lifeboats from *Titanic* to head toward the lights of the unknown ship for rescue.

During the next several hours, the crew of the *Titanic* shot off eight rockets, which were clearly visible to the officers and men on *Californian*'s bridge. At least one officer notified Captain Lord, who was by then in his sea cabin. The captain did not come to the bridge to see for himself, but did ask the color of the rockets. Another crewman later notified Lord about additional rockets, but this was ignored.

After 2:30 A.M., the bridge crew of *Californian* believed the other ship had disappeared. They notified Lord around 2:40 of this development. The captain did not come to the bridge or change his orders to remain stationary.

At 4:30, Lord returned to the bridge, where he acknowledged to another officer that he had been told about the rockets. Just as day was dawning at 6:00, Evans returned to duty at the radio set. He quickly received messages from the *Frankfurt* and the *Virginian*, reporting that *Titanic* had struck an iceberg and was sinking. Lord ordered the engines started and headed toward the reported position of *Titanic*. This position was in error, however, it was a location on the far side of the ice field. Lord carefully picked his way through the ice floes. When he arrived at the presumed location, nothing could be seen. The *Californian* then received another radio message, with the correct position of the survivors. The ship headed south and east, rounding the ice field. They soon saw the *Carpathia*, picking up the final survivors of *Titanic*.

Californian resumed her course for Boston later that day. At Lord's direction, the log did not mention the rockets that had been seen during the night. He also asked Stone and Gibson to write up reports of what they had seen and told him. Lord did not plan to turn these reports over to investigators, however. Instead, he prepared a series of reports and charts that showed *Californian*'s actions and positions, with an eye to justifying his actions, or the lack thereof. After the *Californian* docked in Boston, a member of the crew told a local

newspaper about seeing distress rockets and the failure to act. The news caused a sensation. When reporters asked Lord about the incident and *Californian*'s location when *Titanic* was sinking, he responded by saying the location was a "state secret."

Lord was eventually called to testify at both the American and British inquiries into the disaster. His story changed during the inquiries. When questioned by congressional investigators, Lord admitted to being told of one rocket. At the later British inquiry, he confessed that eight rockets had been seen, and he was told three times about them. Both inquiries found that Lord had failed to respond properly to obvious distress signals. He soon lost his job with the Leyland Line. Although Lord was quickly hired by another shipping line, he spent most of the remainder of his life trying to justify his actions on April 14.

Supporting Proof of Lord's Failure to Render Assistance to *Titanic*

Lord's Attitude toward the Unidentified Ship

Lord reported seeing what he believed to be a masthead light of another ship at around 10:30 P.M. Groves had assured Lord that the light was only a star. Lord, perhaps unusually, agreed. It was only after taking his brief tour of *Californian* that Lord once again noticed a light to the southeast. It was then almost 11:00. At that point, Lord asked the radio operator if he had any information on which ships were nearby. Cyril Evans, the radio operator, reported that he had monitored messages from *Titanic*. Evans knew *Titanic* was nearby because of the strength of her signals. At the time, radio communication was through coded signals, similar to that used in telegraphs. An operator had to use a key apparatus that used short dots and long dashes in place of letters or words. Direct voice communication was not possible. An experienced operator could tell how far away a transmitter was by the strength of the signal. The closer a transmitter, the stronger and more powerful the signal was. *Titanic* had a powerful transmitter, so when she was transmitting nearby, Evans could hardly send his own messages out.

At around 11:30 P.M., Lord returned to the bridge. At that time, *Californian* was drifting slowly. She had turned so that her bow was pointing east-northeast. James Gibson, a 24-year-old apprentice, reported that the light belonged to another ship, which was visible to the southeast. At the time, he estimated the ships were no more than six miles apart. The other ship had apparently stopped. Lord told Gibson to signal to the other steamer with the Morse lamp. The Morse lamp used a powerful light and shutters, to send signals with a combination of long and short flashes of light. On a clear night, like that of April 14, the Morse lamp should have been visible for up to 10 miles. Gibson signaled for several minutes. Lord could see a flickering light on the other ship and believed its crew was returning the signal. The light did not make out a recognizable message, however.

Lord apparently decided the other vessel did not see their signal, although he could certainly see their light. Shortly afterward, he noted a green light on board the other ship. A green light was displayed on all ships on the starboard, or right, side. The light helped other vessels to determine the position and heading of a ship seen at night. Lord realized the other vessel was pointing northwest. Crewmen on the *Californian* later noticed they could see the red light on the port, or left hand, side as well, indicating the other ship was pointed directly at them. Lord went to bed soon afterward.

The *Californian*'s captain displayed a certain lack of urgency about identifying the other ship he could see. The evidence indicates that Lord almost certainly was seeing the *Titanic*. Her approach from the east and lack of movement after 11:30 P.M. coincides with events on the *Titanic*. Her collision with the iceberg was at 11:40, and her engines shut down shortly afterward. The light that Lord saw earlier was most likely the light on *Titanic*'s main mast. From the bridge of the *Californian*, Lord's radius of vision was about six miles. Anything that rose above the horizon could be seen much farther away. *Titanic*'s main mast rose 150 feet into the air, making it visible for up to 16 miles. Lord could have seen it approaching from the east, and Groves may have mistaken it for a star rising over the horizon. An hour later, the *Titanic* would have been much closer to *Californian*, even if its course was somewhat to the south. The green light is another indication that Lord could see the *Titanic*. When she struck the iceberg, the *Titanic* was sailing west-southwest. Her attempts to avoid the iceberg left her bow pointing toward the northwest.

Lord was concerned enough about the strange vessel to the south to issue orders for the bridge crew to watch it and to notify him if it came closer. He naturally wanted to avoid a collision in the dark and was well aware that the other ship was apparently drifting, just as the *Californian* was. Even if his first attempts at signaling with the Morse lamp failed to get a response, the captain could have ordered his crew to try again later. Lord could also have roused his radio operator to try to determine the other ship's identity. Evans had just gone to bed and Lord did not try to wake him. Thanks to his conversation with Evans earlier, Lord was aware that the *Titanic* was relatively close by.

Failure to Respond to Signal Rockets

One of the most important charges against Lord was that he failed to respond to rockets sent off by crewmembers of the *Titanic*. Fourth Officer Joseph Boxhall of *Titanic* testified that he began shooting off distress rockets about an hour after *Titanic* struck the iceberg. Following standard procedures, he shot them off at regular intervals. The rockets went several hundred feet into the air with a white light before exploding into stars.

The rockets were noticed by the crew on the *Californian*. Ernest Gill, an assistant engineer who had been in the engine room, came off watch at midnight.
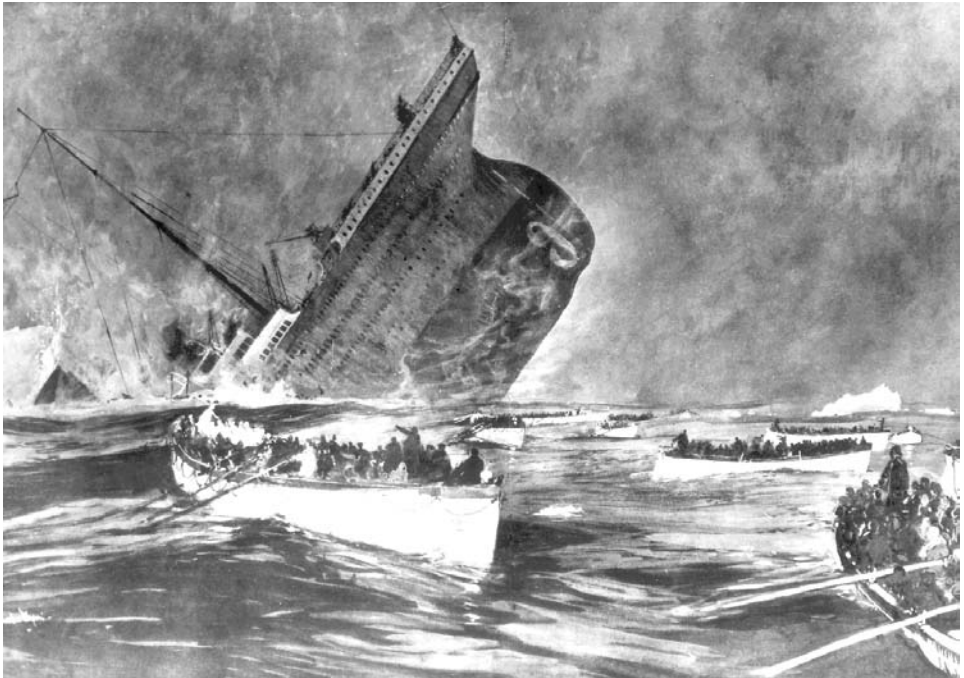
A depiction of the *Titanic*, sinking during her maiden voyage in April 1912. (The Illustrated London News Picture Library)

He observed a large steamship within 10 miles away and clearly observed its lights. About an hour later, he came back on deck to smoke and saw the other ship shoot off a white rocket. He immediately believed it was a distress signal. Gill did not report it to the bridge because he assumed they were watching the other vessel and could see the rocket as well as he did. Stone, who was standing watch on the *Californian*'s bridge, noticed one rocket at 12:45. He observed several more rockets before calling down to Lord in his cabin at 1:15. Stone reported that the strange vessel to the south was firing rockets. Lord asked about the color of the rockets and whether they fit the pattern of rockets used by any shipping line. When Stone replied they were white and did not have any company colors, Lord went back to sleep.

In 1912, the International Rules of the Road issued to ships and commanders clearly set out how ships could signal when they were in distress. The first way was to sound an explosive signal, such as a gun, at regular intervals, such as a minute apart. Another accepted method at night was to shoot off rockets of any color, one at a time at regular intervals. Experienced sailors in almost any location could hardly have viewed the rockets as anything other than a distress signal.

Defenders of Lord have sometimes cited the idea that distress rockets were supposed to be green or red, or some other color. They tried to indicate that white rockets were not generally regarded as distress signals. Some suggested

that the crew believed the rockets were part of the entertainment being staged for the passengers of the unknown vessel.

The International Rules of the Road, however, did not specify that distress rockets had to be any particular color or colors. Some combinations of colored rockets and lights were used in specific locations to identify the company to which a ship belonged. The Merchant Shipping Act of 1894, which regulated British merchant ships in 1912, permitted companies to use specific signals. The act authorized the Board of Trade to produce a pamphlet outlining company signals. These combinations of colored rockets or lights were unique to individual companies. The Leyland Line, for example, was authorized to burn three red lights in quick succession to identify its ships. A key provision in the regulation, however, was that these signals were to be used only in certain areas of the sea and only for identification purposes. The North Atlantic was not an authorized location for company identification signals. In its preface, the pamphlet stated that if signals were used in any other place, they might be considered signals of distress. Captains who observed signals, such as rockets being launched at regular intervals, would take them as distress signals. They were expected to react accordingly.

Between 12:45 and 1:15, Stone observed five rockets from the other ship. He later testified that he understood these rockets to be distress signals. He did not explain why he failed to act or to inform Lord that he believed the other ship was in trouble. Stone had taken his examination for a certificate as first mate only four months earlier, and this information was tested on the examination. He only noted that the strange vessel had disappeared at 2:40. He believed it had sailed away.

A question also exists about whether the *Californian*'s crew was close enough to hear the rockets' explosions, another distress signal. Because the night was clear and calm, the sound of the explosions might have been heard over the 10 miles or so that separated the two ships. No crewmen on *Californian* reported hearing any explosions. Although making a regular sound or constant noise was considered a signal of distress by the International Rules of the Road, sound was not considered a primary indicator of problems. Although Lord's defenders made much of the failure to hear the rockets' explosions, this lack was not an excuse to ignore the rockets as a sign that another ship was in trouble.

Failure to Wake Californian's Radio Operator

Another aspect for which Lord was criticized was the failure to use the radio to try to identify the unknown ship and to learn about its distress. Lord was familiar with the capabilities of the new technology. Although this was only the second voyage on which *Californian* carried a Marconi set, Lord had authorized its frequent use and willingly acted upon the information received via the radio. During the past few days, *Californian* received a number of messages from

other vessels about sailing conditions. They were particularly useful when *Californian* entered the region where icebergs and other ice could be encountered. Earlier on April 14, Lord had asked Evans which ships were nearby when he thought he saw another ship approaching.

When attempts to contact the strange ship by Morse lamp failed, Lord did not try to use the wireless. When he realized Evans had gone to bed just before 11:45 P.M., Lord did not rouse him. Although at that time no apparent emergency existed, things had obviously changed just an hour later. When the rockets were reported to Lord, he could have ordered Evans to return to duty and monitor the radio frequencies. After striking the iceberg, the *Titanic* was continuously sending out radio messages, asking for help. By ignoring the possibility that the rockets were a call for help, Lord failed to use all the tools at his disposal to verify whether an emergency existed. If Evans had monitored the radio traffic for just a few minutes, he would have been able to tell Lord that the *Titanic* was sinking just a few miles away.

Failure to Take a Direct Route to the Titanic's Location

Lord arose at 4:30 A.M. on April 15. He made his rounds of the ship to check the conditions and the ice around it. By 5:00, he was on the bridge. Lord sent an officer to rouse Evans and check on the rockets that had been reported during the night. The officer quickly returned with news that the *Titanic* had sunk. Lord sent him back to get a location. The first location was 41° 46' N, 50° 14' W. This location would have been nearly the same as the ship that was observed the night before. Lord studied the position, then sent his officer back to check on whether other ships had the same location of the sinking. A different position was then received, on the other side of the ice field that had stopped *Californian* the night before.

Lord cautiously picked his way through the ice in a southwesterly direction. By 6:20 A.M., they were clear of the thickest ice and Lord ordered full steam ahead. The *Californian* was making about 13 knots. The *Californian* passed the reported position at 7:30 and saw two other ships, but no sign of a wreck. They continued south until Evans received a message from the *Carpathia* that it was picking up the *Titanic*'s survivors. Lord turned his ship back to the east and pushed through the ice field again. By the time they approached the *Carpathia* at 8:30, the rescue operation was completed. The *Californian* circled the area looking for survivors for about three hours, then resumed its course to Boston.

By the time Lord arose at 4:30, the *Titanic* had been sunk for two hours. When he finally received information about radio messages that the *Titanic* had struck an iceberg, he prepared to rescue survivors. However, Lord failed to act on the first position reported. This failure seems odd because the position would have been close to the ship observed the night before, firing rockets. Instead,

Lord waited until he had a different position, one farther away and on the other side of the ice field. That position would have excused him from rendering assistance because of the distance and the ice. Sailing to that position prevented the *Californian* from arriving at the true location of the *Titanic*'s sinking until it was too later to render any practical assistance.

Lord's Attempt to Cover Up

Californian reached Boston after the *Carpathia* had docked in New York, and the survivors of the *Titanic* had time to tell their stories. A congressional inquiry into the loss of the liner was scheduled, but Lord did not tell anyone about what he and the crew had seen on April 14–15. He had asked Groves and Stone to write up a report of the night, but did not share the report with anyone. Lord also made certain the log did not include any mention of the rockets.

Lord and Evans were subpoenaed to testify before the inquiry, but mostly about the conditions in the area. Other ships had recorded the *Californian*'s message about ice, and investigators wanted some background information. When reporters asked Lord about what had happened, he told them that the *Californian* had been too far away from the *Titanic* to help. Lord's comments angered Gill, who gave a signed affidavit to the *Boston American*. In it, Gill reported that eight rockets had been seen from *Californian* and that the officers on duty had reported them two or three times to the captain. Lord did not see the newspaper story before he testified, although Senator William Smith was aware of it. He was able to surprise Lord with questions about the rockets during Lord's testimony. Lord was forced to admit that he had been told of at least one rocket and to admit that the ship's log did not include mention of rockets that could have been distress signals from the *Titanic*.

Lord's attempt to dodge responsibility drew criticism from the press. In the British inquiry that followed the American one, Lord was again asked what was seen from the *Californian* on April 14–15. He was more forthcoming, but again tried to deny any failure to come to the *Titanic*'s aid. Both inquiries found Lord negligent for his failure to offer assistance.

Lord's deception about what had been seen from the *Californian*'s bridge led to criticism in the press. His attempts to discredit his own crewmen in public, then being forced to admit the validity of the crewmen's testimony, caused Lord to lose credibility. For many members of the public, Lord's attempt to hide something made them believe he had more to hide.

Historical Interpretations Supporting Lord's Guilt in Failing to Render Assistance

The question of whether or not the *Californian* could have given assistance to the *Titanic* and rescued many of the passengers that were lost has attracted great

attention. Both the British and American inquiries condemned him. Other members of the *Californian*'s crew wrote criticisms of him, including calling him a "thousand-fold murderer." Others quickly spoke up in his defense. Soon after the loss of the *Titanic*, British solicitor A. M. Foweraker published several articles defending Lord. Lord's defenders became known as "Lordites." They used different arguments, including one that an unidentified mystery ship actually was between the *Titanic* and the *Californian* and was seen by both ships. They also were accused of suppressing unfavorable evidence and playing up comments most supportive of Lord.

After a while, the furor over the *Titanic* died down. The outbreak of World War I only two years later helped push the disaster out of the public consciousness. The controversy over what Lord saw and did or did not do resurfaced in 1955. That year, Walter Lord published *A Night to Remember*, a popular history of the *Titanic* disaster. The book and the movie that was later made from it were critical of Lord and his failure to act. Lord, who had kept a low profile, was spurred to try to clear his name. He filed a series of appeals with the Mercantile Marine Service Association to review the facts and establish whether the *Californian* could have seen the distress rockets and saved the *Titanic*'s passengers.

A series of books on both sides of the argument followed. Those that were critical of Lord included the following. Roy Anderson, in *White Star* (1964), covered the history of the White Star Line. Anderson blamed Lord for not coming to the assistance of the *Titanic*. More important, Leslie Reade and Edward de Groot wrote *The Ship that Stood Still: The Californian and Her Mysterious Role in the Titanic Disaster* (1993). Reade used research by Leslie Harrison, who supported Lord, to discredit the captain's claims. Harrison tried to block publication of *The Ship that Stood Still*, but was unable to prevent an edited version from appearing after Reade's death.

Conclusion

In 1992, Lord's request for a review was finally granted. Because the *Titanic*'s wreck had finally been located, her position at sinking was established by a review board. According to the review board, the distance between the *Californian* and *Titanic* was 13.5 miles, close enough for distress signals to have been seen. The board believed the *Californian*'s crew had seen the distress signals and Lord had failed to act according to established rules. Whether he could have saved all the passengers and crew was debatable. Lord's failure was again documented, 80 years after the fact.

References and Further Reading

Carrothers, John C. *Lord of the Californian*. Annapolis, MD: U.S. Naval Institute, 1968.

Eaton, John P., and Charles A. Haas. *Titanic, Triumph and Tragedy*. New York: Norton, 1995.

Great Britain. Marine Accident Investigation Branch. *RMS Titanic: Reappraisal of Evidence Relating to SS Californian*. London: H.M.S.O., 1992.

Harrison, Leslie. *A Titanic Myth: The Californian Incident*. London: W. Kimber, 1986.

Padfield, Peter. *The Titanic and the Californian*. New York: John Day, 1965.

Rasor, Eugene L. *The Titanic: Historiography and Annotated Biblio*graphy. Westport, CT: Greenwood, 2001.

Reade, Leslie, and Edward P. de Groot. *The Ship that Stood Still: The Californian and Her Mysterious Role in the Titanic Disaster*. New York: Norton, 1993.

Streissguth, Thomas. *The Sinking of the Titanic*. San Diego, CA: Greenhaven, 2002.

CON

The RMS *Titanic* tragedy has permeated popular culture since that iconic ocean liner struck an iceberg on its maiden voyage and sank during the early hours of April 15, 1912, resulting in the deaths of approximately 1,500 people. That disaster has been one of the most frequently examined historical maritime incidents. Newspapers, magazines, books, films, television programs, and other media have selectively portrayed *Titanic* crew members and passengers of various classes as heroes, victims, or villains. Depictions often represented opinions and stereotypes describing people who had perished that night and thus could not defend their actions or correct inaccuracies. Myth and fact sparred when some authors' exaggerated, ignored, or fictionalized events in an attempt to intensify drama and romanticize characters and their reactions, usually cast as courageous, to abrupt life-and-death situations and sacrifices they elected to make. Some writers purposefully distorted aspects relevant to the sinking in order to criticize specific individuals and absolve others for the horrific losses suffered that bitterly cold April morning.

Accounts also featured people associated with the *Titanic* who were not aboard that ship. Captain Stanley Lord, who was master on the British steamer SS *Californian*, is perhaps one of the most vilified and misunderstood people who became embroiled in the *Titanic*'s misfortunes. Two official government inquiries in 1912 stated Lord had seen distress signals from the *Californian* and been able to reach the *Titanic* to save survivors but had refused. Discussion whether Lord neglected to assist the *Titanic* represented one of the most controversial *Titanic* topics in the press at the time of the disaster and has been a recurring issue throughout the century since that liner sank. Two sides, the Lordites who believe

the captain was innocent of negligence and the anti-Lordites who think he was guilty, have debated this issue for decades in various forums and groups such as the Titanic Historical Society, expanding discussion globally on Internet message boards and Titanic Web sites.

Captain Stanley Lord and the *Californian*

When the *Titanic* sank, Captain Stanley Lord was an accomplished seaman. A native of Bolton in Lancashire, England, Lord was born to James and Charlotte Lord on September 13, 1877. Lord and his five older brothers grew up in the Bolton community, where one of his brothers was classmates with Arthur Rostron, who later was captain of the *Carpathia* when that ship's crew rescued *Titanic* victims. As a child, Lord yearned to join the Mercantile Marine so he could see the world. No one in his immediate family had been employed in naval pursuits, and young Lord had no experiences with that profession except through sea stories he had read and heard. Although Lord's parents disliked the idea of their son pursuing a naval career, Lord convinced his father to sign an indenture with Lord on May 7, 1891, at Liverpool, permitting Lord to join J. B. Walmsley & Company for six years of service, beginning on the *Naiad*, a barque belonging to his employers. Lord excelled at seamanship and earned a second mate's certificate and was hired as the *Lurlei*'s second officer.

By 1897, Lord began employment with the West India and Pacific Steam Navigation Company and joined the Mercantile Marine Service Association (MMSA), a professional membership he retained throughout his career. Three years later, Frederick Leyland & Company, Ltd. (often referred to as the Leyland Line) gained control of that navigation company, and Lord agreed to work for them. He secured master certification during February 1901 then qualified for an extra master's certificate in the spring of that year. Lord served on routes from Europe to North and South America and Africa. He participated in training events, including a lifeboat exercise involving 1,000 sailors at an Essex beach in September 1904. Lord's achievements surpassed those of many officers serving on British vessels. His peers admired his capabilities, work ethic, and conduct, describing Lord as responsible and methodical.

Lord first served as a commander during 1906. He was master of the *Antillian*, *Louisianian*, and *William Cliff* prior to service on the *Californian*. Caledon Shipbuilding and Engineering Company, Ltd. built the *Californian* at Dundee, Scotland, in 1901. Designed to transport cotton and passengers, the *Californian*, owned by the Leyland Line, weighed 6,223 tons and was approximately 447 feet long, almost 54 feet wide, and 35 feet deep. That steamer completed its maiden trip in 1902. Starting in 1911, Lord served as the *Californian*'s fourth captain, guiding that cargo steamer filled with cotton bales between New Orleans, Louisiana, and Le Havre, France.

April 14, 1912

On April 5, 1912, Lord and a crew of approximately 50 men departed from Victoria Docks at London, England, in the *Californian*. No passengers were aboard for that trip. As the *Californian* moved westward toward its destination, Boston, Massachusetts, ship crews also traveling in that region sent wireless communications the *Californian* received that stated ice was present on the North Atlantic Ocean shipping routes.

Lord stood on the *Californian*'s bridge to survey the sea on the evening of April 14. He saw three icebergs at 6:30 that night and had his ship's wireless operator, Cyril Furmstone Evans, warn the Leyland Line

Jack Phillips, wireless operator on the *Titanic*, 1912. (The National Archives/Heritage-Images/StockphotoPro)

vessel, *Antillian*, regarding the ice hazard. In that message, Evans included the *Californian*'s location, 42° 5′ N, 49° 9′ W, at that time. Later, Lord detected ice

---

Jack Phillips and the *Titanic*'s Wireless

The wireless operator on the RMS *Titanic* who took the warning of icebergs from the SS *Californian*, but then admonished the operator on the *Californian* for interrupting his duties, was John George "Jack" Phillips. He was born on April 11, 1887, at Farncombe, Surrey, southwest of London. His father, George Alfred Phillips, was a draper, originally from Wiltshire, and it was expected that Jack might go into the family business as did one of his older sisters. The young Jack Phillips went to school at Godalming and then started working at the local post office. It was there that he learned about telegraphy and managed to find work for the Marconi Company in 1906.

Phillips went to work for the White Star Line and was appointed to the *Teutonic*, and then to the *Campania*, the *Corsica*, the *Victorian*, the *Pretorian*, the *Lusitania*, and then the *Mauretania*, before working at the Marconi Station near Cliften, Ireland. He then went back to sea on the *Adriatic* and the *Oceanic*. During this time he sent out thousands of messengers for passengers—something extremely lucrative for the shipping line. It was as a result of this that he was said to have been annoyed at the interruption by the message sent by the *Californian*.

---

in his steamer's path and ordered his crew to cut power to the engines at 10:21 P.M. The *Californian* stopped next to the eastern side of an immense field of ice approximately 30 miles in dimension. Lord had never maneuvered vessels near field ice and decided to delay the *Californian*'s movement until the next morning when sunlight would aid him to evaluate the situation. His decision to stop was a typical response many sailors chose when confronted with ice fields. Lord used his compass to determine the *Californian*'s position was 42° 5′ N, 50° 7′ W when it stopped.

To alert vessels concerning the ice field, Lord told Evans at approximately 11:00 to issue a radio message that stated the *Californian* had ceased movement because of ice. The *Titanic*'s wireless operator, Jack Phillips, responded to the *Californian*'s warning by tapping out an abrupt message stating that he was busy and to leave him alone. On the *Californian*, Evans, who had been working since early that morning, left his wireless equipment after 11:30 and fell asleep. About that time, Lord and several crew members saw a ship stop about five miles south-southeast of the *Californian* and decided that it was another steamer. They used Morse lamps, which rotated 360 degrees and could be seen at least 16 miles away, especially in a clear sky like they experienced that night, to signal the vessel, but it did not respond.

About 20 minutes before midnight, the *Titanic* hit an iceberg and began filling with seawater. Phillips issued an SOS message, stating that the *Titanic* was in distress at 41° 46′ N, 50° 14′ W. As water filled the *Titanic*'s lower levels, passengers and crew put on flotation vests and began assembling on decks near lifeboats, which there were not enough of for everyone aboard. In the distance, about five miles away, many people on the *Titanic* saw ship lights. Crew members began firing distress signals approximately 200 feet high in the sky and using similar lamps like the *Californian* had to send Morse code messages, hoping that unidentified ship and any others nearby would assist to save *Titanic* passengers. The mystery ship approached the *Titanic* but then departed from sight without helping.

April 15, 1912

Meanwhile, by 12:40 A.M. Lord, tired by a full day's work and wearing his clothes and boots, napped in the chartroom. He had ordered the second officer, Herbert Stone, to wake him if needed. Stone used a speaking tube around 1:15 to tell Lord the steamer they had seen was moving southwest. Stone reported he had seen a rocket low in the sky near the steamer and used Morse lamps to signal it. Because he did not receive any responses before the steamer vanished from view in the southwest, Stone decided that the rocket had not been urgent. Lord resumed sleeping, feeling assured the steamer did not need assistance.

Californian officers and crew members who were awake during the night observed rockets at a distance in the sky southeast of their ship's position. They

thought they were seeing signals between ships from the same shipping line or fireworks for a party, not distress flares. The flares were low on the horizon, not high like distress signals, and the crew did not hear explosions associated with emergency rockets. Later, Lord stressed that if he and his men had been aware of the *Titanic*'s distress calls and signals, they would have attempted to move through the ice in the dark and headed to the *Titanic*.

By dawn on April 15, Evans resumed monitoring radio operations, reporting the *Californian*'s coordinates at 5:15 A.M. Receiving messages from the ships *Frankfurt* then *Virginian*, the *Californian* crew became aware that the *Titanic* had sunk. About 6:00 Lord ordered his crew to maneuver through the ice field until they could proceed at full speed to the disaster site according to coordinates in the *Titanic*'s distress messages. When the *Californian* reached those coordinates, Lord saw no signs of the *Titanic*.

Hoping to find that ship, Lord guided the *Californian* south until he detected the *Carpathia* and its crew led by Captain Rostron surveying the *Titanic*'s actual sinking site where debris such as deck chairs floated on the surface. Lord estimated coordinates where the *Titanic* sank were 41° 33′ N, 50° 1′ W, approximately 33 miles from where the *Californian* had been stopped. The *Californian* arrived after the *Carpathia*'s crew had secured *Titanic* survivors from lifeboats. Lord and his crew searched the area for anyone the rescuers might have missed but did not find anyone. By 11:20 A.M., Lord ordered his crew to head toward Boston where they docked on April 19.

Official Inquiries

Soon after his arrival in Massachusetts, Lord traveled to Washington, D.C., when a U.S. Senate Commerce Committee chaired by Senator William Alden Smith requested Lord to be a witness at a government hearing investigating why the *Titanic* sank. On April 26, 1912, the committee questioned Lord and two *Californian* crew members, Evans and boilerman Ernest Gill, who had received $500 from the *Boston American* for his account. Gill claimed Lord had seen a moving passenger ship's port side light, implying that vessel had been the *Titanic*, and distress flares but refused to help. Lord denied Gill's allegations, but the senators accepted Gill's account as the truth.

In its May 28, 1912, report, Smith and committee members asserted their belief that the *Californian* had been closer to the *Titanic* than Lord indicated and concluded the ship lights the *Titanic* survivors had seen had been the *Californian*'s lights. Committee members insisted that if the *Californian* crew had taken action when they saw the flares they could have rescued everyone on the *Titanic*. The committee rebuked Lord for his inaction and negligence, establishing a precedent for further scrutiny of Lord and preserving its harsh judgments in public records, which omitted many crucial facts.

Lord went to Liverpool to give a statement to Leyland Line Marine Superintendent Captain Fry and became aware that Charles Victor Groves, the *Californian*'s third officer, also said the ship the *Californian* crew saw on April 14–15 was a passenger vessel, suggesting it was the *Titanic*, and that he had seen a port side light and watched the deck lights go out prior to midnight. Lord clarified that was the first time he had heard Groves make that claim.

He next appeared as a witness to assist the British Board of Trade inquiry led by John Charles Bigham, formally known as Lord Mersey, examining why the *Titanic* sank. During those hearings in London, Lord testified on May 14 and was never designated as a defendant or tried for any crimes. He was represented by his employer's counselor, C. Robertson Dunlop, who did not interact with Lord outside the inquiry to comprehend Lord's navigational evidence. On May 15, Lord was dismayed when Lord Mersey urged Groves to agree with him that the *Californian* crew had seen the *Titanic* nearby. Lord returned to Liverpool that night because he was scheduled to sail three days later but was told by the marine superintendent to retrieve his belongings from the ship because he would not be going to sea.

In June, a month after Lord concluded his testimony and was not present to defend himself, British investigators questioned why Lord and the *Californian* had not reached the disaster site until hours after the *Titanic* sank. Like the U.S. investigators, the British officials decided that the unidentified ship the survivors stated they had seen near the *Titanic* was the *Californian*, despite Lord's assertions that his ship had been stationary throughout the night.

At the British inquiry, officials accepted the *Titanic*'s position in messages as valid and dismissed other evidence as inaccurate. Leyland & Company presented the British investigators a communication sent from the *Californian*, noting its position, to the *Antillian* before the *Titanic* sank, but this evidence was discounted. Investigators believed witnesses who lacked credibility and did not investigate information provided by more reliable witnesses such as Stone, who stated the ship he had seen near the *Californian* was moving normally and not in distress.

A July 30, 1912, report written by Lord Mersey stated the *Californian* had been within 8 to 10 miles from the *Titanic*, witnessed distress signals, and could have safely navigated to the disaster site and rescued *Titanic* survivors. The White Star Line supported the government's findings, which focused on Lord and not the line's negligence to provide sufficient lifeboats. Despite its harsh charges regarding Lord, the Board of Trade did not seek his prosecution and permitted Lord to retain his certification.

Persecution of Captain Lord

Lord initially became aware of allegations accusing him of negligence in newspaper accounts of the inquiry's report. He contacted the MMSA for assistance

and believed those accusations would be corrected if he presented evidence. On August 10, 1912, Lord wrote the Board of Trade about the hearing, specifying facts about the *Californian*'s location. He stated the investigators had not consulted logbooks or sought other ships in the area. Lord emphasized the distance between the *Californian* and *Titanic* would have delayed any rescue response if he had been aware of the sinking and the *Californian* would have been unable to reach *Titanic* survivors before the *Carpathia* did. He asked the board to reconsider its judgment against him so people would know he was not guilty of any wrongdoing regarding the *Titanic*.

The British inquiry report and public disdain ended Lord's career with Leyland & Company when its directors pressured for Lord's dismissal. Lord surrendered his commission on August 13. Leyland executives, who believed in Lord's innocence, stated he might be reinstated if the Board of Trade reversed its decision against him. Lord and MMSA officials unsuccessfully sought for the Board of Trade to reconsider Lord's evidence. The MMSA *Reporter* printed a letter from Lord, presenting information omitted from the inquiry and countering false charges regarding his conduct as *Californian* captain. Lord released a statement, declaring the *Californian* was not the ship seen from the *Titanic* and that he had not acted negligently. He stressed the improbability of implications he could have rescued the *Titanic*'s survivors and insisted he was a qualified commander who served with integrity and professionalism.

Lord's attempts to prove his innocence were mostly ignored. Public opinion blamed Lord for abandoning *Titanic* survivors. Newspapers and magazines criticized Lord, printing stories with incorrect information and attacking his character by labeling him a drunkard, liar, and coward. During the hearing, Attorney General Sir Rufus Isaacs had hinted Lord had been inebriated while commanding the *Californian*. Lord, who insisted he was always sober, offered additional evidence in letters from colleagues, including *Carpathia* Captain Rostron who told Lord that, in his opinion, Lord Mersey had acted unprofessionally. W. H. Baker sent Lord a letter saying he had heard crew members of the *Mount Temple* state that their ship had been near the *Titanic* and left without helping. The Board of Trade did not secure testimony and facts to verify any of these people's information. Lord contacted A. H. Gill, a Parliament member from Bolton, and went to the House of Commons on October 23, 1912, but was frustrated in his efforts to obtain justice.

Professional Survival

By February 1913, Lord resumed commanding ships, sailing for the Nitrate Producers Steamship Company. Frank Strachan, the Leyland Line's United States agent, had recommended Lord to that company's owner, John Latta. That year, Lord's dilemma attracted supporters such as Albert Moulton Foweraker, an attorney and yachting enthusiast from Cornwall. Foweraker wrote articles

titled "A Miscarriage of Justice," which *Nautical Magazine* printed in several 1913 issues. He elaborated why evidence proved Lord could not have rescued the *Titanic's* survivors. Prominent journalist Filson Young endorsed Lord in major periodicals, stating people's biases, not facts, had unjustly censured Lord.

In August 1913, Lord continued efforts to secure a new hearing of his evidence and referred to Foweraker's articles in his request to the Board of Trade. Lord commented that he suffered daily because of the false allegations and desired to prove his innocence to restore his reputation. He stated that if the inquiry had attempted to identify steamers near the *Titanic* they would have realized which one was the mystery ship and not have targeted him.

Lord commanded Nitrate Producers Steamship Company vessels for fourteen years, serving during World War I by ferrying horses and eluding submarine attacks on his ship. He retired in March 1927 when cataracts impaired his vision. During the next three decades, Lord did not dwell on the unfair incriminations he had endured in 1912.

Renewed Publicity

The movie *A Night to Remember*, released in 1958, and Walter Lord's book of the same title which it was based on, provoked Lord's outrage that long-forgotten accusations against him had resurfaced. He was incredulous that the film portrayed the *Californian* as being near the *Titanic* when it sank and that Lord had refused to help. Stating he wanted to clear his name, Lord consulted MMSA officials and asserted that government officials had wrongly maligned him in 1912.

That association's general secretary Leslie Harrison, who was born in July 1912, consulted Lord's evidence to familiarize himself with facts. Assured of Lord's credibility, Harrison defended Lord to the press, which had printed numerous articles critical of Lord based on what the film claimed had happened. Harrison wrote the film's producer, book's publisher, and Walter Lord, requesting they recognize Lord's evidence, but those individuals stated they considered the inquiry report factual. He complained to theater managers to stop showing *A Night to Remember*, but the film continued to be distributed commercially and to schools. Lord considered filing libel lawsuits, but poor health limited his activities. The stress weakened Lord, who had been widowed when his wife, Mabel, died in 1957.

Lord had a detailed affidavit of his experiences the night *Titanic* sunk notarized on June 25, 1959, hoping that the statement might be used eventually to exonerate him. Again, he emphasized the *Californian* had been stopped in an ice field from before midnight through dawn on April 14–15, 1912, and could not have been the steamer seen near the *Titanic*. Lord gathered documentation, such as his logbooks and charts, which the 1912 U.S. and British inquiries had not allowed him to submit as evidence, and presented them to Harrison.

Posthumous Deliberations

Lord's January 24, 1962, death did not end the controversy regarding his command of the *Californian* in April 1912. His son, Stanley Tutton Lord, who had been born four years prior to the *Titanic* tragedy, defended his father's honor for the next 30 years until he died in 1994. Harrison also continued seeking public vindication for Lord. In early September 1963, Harrison contacted British Minister of Transport Ernest Marples, requesting he reexamine evidence from the 1912 inquiry and consider information that had became available since then in an effort to exonerate Lord posthumously.

For example, Arthur Swinson and Stanley Williamson wrote a British Broadcasting Corporation script, "The Other Ship," which discussed how Henrik Naess admitted to Norway's consul in Iceland that the *Samson*, a Norwegian vessel on which he had served as first officer in April 1912, had been near the *Titanic*. Naess stated that he and the *Samson*'s lookout had observed ship lights and rockets on April 14–15, 1912, and that those lights vanished. Without a wireless radio to receive messages, Naess said the *Samson* had quickly departed because the crew misinterpreted the rockets as possibly being from a ship with revenue personnel who had seen the Norwegians poaching seals. Naess only realized they had been mistaken when he learned about the *Titanic* after returning to port and looked at that date in the *Samson*'s logbook. His admission was classified and first publicly released in April 1962, several months after Lord's death.

By February 1965, the MMSA, frustrated because Marples had rejected Harrison's 1963 petition for a hearing, again prepared to request a new investigation of Lord to prove he had not acted dishonorably. The group made another attempt in March 1968 for the Board of Trade to review unexamined evidence, including an affidavit by science teacher Lawrence Beesley, a *Titanic* survivor, who stated that the ship he saw near the *Titanic* was not the *Californian*. Beesley also said he saw all the *Titanic*'s distress flares fired before the lifeboat he was in floated away from the *Titanic* at approximately 1:30 A.M. Crew members on the *Californian* had seen flares as late as 2 A.M.

Historical Perspective and Analysis

Government refusal to consider new evidence provoked several authors, many of them with seafaring experience, to present historical narratives addressing Lord and the *Californian* based on available primary resources. Peter Padfield, who had been a Merchant Navy officer, defended Lord in his 1966 book, *The Titanic and the Californian*, examining how the British inquiry failed Lord by focusing on unreliable witnesses, ignoring facts, misunderstanding coincidences, and distorting information. Padfield emphasized that Lord Mersey had purposefully concentrated on Lord to prevent criticism of British shipping from rivals,

particularly Germans, and to obscure the Board of Trade's insufficient enforcement of emergency regulations. By holding Lord accountable for failing to rescue survivors, Padfield suggested the Board of Trade could avoid scrutiny of *Titanic* crew errors.

In 1968 John C. Carrothers, who served as an engineer for several shipping lines, reassessed the 1912 U.S. and British inquiries in a *U.S. Naval Institute Proceedings* article. Carrothers had previously published an April 1962 essay in that journal that was critical of Lord based on those inquiries' reports, then received a letter from Harrison that alerted him to evidence omitted from or misconstrued during the hearings. Carrothers's 1968 article emphasized facts that proved Lord had not been guilty of ignoring the *Titanic*. Carrothers explained how several witnesses' testimony was flawed, noted Lord Mersey's inexperience with sea vessels and navigation, and suggested reasons the Board of Trade focused on Lord, including an attempt to prevent negative publicity about prominent people, including White Star Line executive J. Bruce Ismay, surviving in lifeboats while so many passengers drowned or died from hypothermia.

The 1985 discovery of the sunken *Titanic* revealed that site deviated south by approximately 3 miles and east by 13 miles from the location specified in emergency messages Phillips sent. Using Lord's materials, Harrison wrote the book, *A Titanic Myth: The Californian Incident*, which was initially published in 1986 and reissued since with varying titles into the 1990s. Harrison examined Board of Trade files the Public Record Office made available to researchers. Among his many arguments supporting Lord's innocence, Harrison emphasized that both the U.S. and British inquiries relied on inaccurate locations for the *Californian* and *Titanic*. Unlike Padfield, Harrison did not think the Board of Trade inquiry intentionally tried to frame Lord for not rescuing *Titanic* survivors. He noted what he perceived as the incompetence of the people directing both inquiries and stressed that Lord Mersey and his legal peers misused their power because they had been confident of Lord's guilt and had no motivation to investigate if any other vessels might have been the mystery ship.

In 1990, the British government agreed to Transport Secretary Cecil Parkinson's request to reexamine the 1912 inquiry. Marine Accident Investigation Branch (MAIB) director Captain Peter Marriott asked James de Coverley, branch deputy chief inspector, to investigate. De Coverly concluded that the *Californian* had been a distance from 17 to 20 miles north of the *Titanic* and Lord could not have witnessed the sinking. Despite this acknowledgment of the ship's position, the MAIB report released on April 2, 1992, stated Lord had observed the *Titanic*'s distress signals. The report admitted Lord and the *Californian* would not have been able to reach the disaster site in time to assist survivors.

Also in 1992 Lieutenant Commander Craig McLean of the National Oceanic and Atmospheric Administration and reporter David L. Eno published "The Case for Captain Lord" in *Naval History*. Eno secured information in

Norway and Iceland regarding Naess and the *Samson* that he stated proved that that vessel had been the mystery ship viewed from the *Titanic*. The authors presented facts about the *Californian*'s location, including a diagram showing how the Earth's curvature would have affected observations of flares from ships at varying distances. McLean and Eno suggested the *Californian* crew had simultaneously seen the *Samson* and flares sent from an unseen ship farther away, probably the *Titanic*. They determined both inquiries had wrongly accused Lord of seeing the sinking *Titanic* and failing to help.

In the early 21st century, Senan Molony, an Irish journalist and *Titanic* historian, examined evidence relevant to Lord and the *Californian* during the *Titanic* disaster. In his comprehensive books (2006, 2008), particularly *Titanic and the Mystery Ship*, Molony discounted allegations that the *Californian* had been the ship *Titanic* survivors stated they had seen near the sinking liner. He explained why various assumptions and statements the inquiries had relied on in 1912 were implausible, illogical, or inconsistent with facts. British reporter Thomas B. Williams (2007) also assessed records to write an account, *Titanic and the Californian*, favoring Lord, promoting his innocence and criticizing the faulty inquiry conclusions regarding the *Californian*'s location and Lord's character.

Conclusion

Most Lordites agree that the 1912 inquiries swiftly found fault with the *Californian* captain based on investigators' review of inadequate and flawed evidence. Those investigators' emotional reactions to the *Titanic* tragedy and impulsiveness to blame someone resulted in the persisting misunderstandings that continued to taint Lord's reputation even after his death. Not considering all relevant evidence, the investigators assumed the *Californian* was the mystery ship *Titanic* survivors saw and quickly decided Lord was guilty without allowing him the opportunity to defend himself and appeal their rushed judgments. The inquiry committees disregarded Lord's statement that the *Californian* had been stopped that night and did not seek information about other ships and captains in the North Atlantic's busy shipping lanes nor evaluate witnesses' testimony completely to determine possible reasons for differing accounts.

Both the investigators and many writers dismissed Lord's insistence that he had seen other steamers that night but not the *Titanic*. The immense size of the brightly illuminated *Titanic* would have easily differentiated it from other ships on transatlantic routes during 1912 because most of those vessels were a maximum of 7,000 tons with those over 20,000 tons being limited to 12 ships. Lord knew the ship he saw several miles from the *Californian* was not the *Titanic* because he had previously seen the *Olympic*, the *Titanic*'s sister ship, from five miles away. The sidelights on the ship Lord and his crew saw indicated that vessel was heading west, not east, like the *Titanic* had been traveling.

Facts regarding the ships' coordinates, which were validated in the late 20th century, proved Lord had been telling the truth. Even if Lord had known about the *Titanic*'s distress, his efforts to reach that vessel would have required several hours for him to remove the *Californian* from the ice field and steam to the coordinates in messages and then continue searching for the *Titanic* when he could not find it at that site. At that time Lord lacked access to such modern technological devices as satellite imagery or sonar, which might have assisted rescue attempts. Neither inquiry addressed whether Lord's vision was a factor or if cataracts may have blurred his vision.

Knowing that he had done nothing wrong on April 14–15, Lord trusted that the inquiries would focus on shipping safety and how to prevent future disasters. His naivete made him vulnerable to politicians who lacked navigational experience and wanted to appease colleagues and public outrage by targeting someone easy to vilify. Their errors became deeply enmeshed in *Titanic* popular culture, which often appropriated and reiterated false information from the inquiries for entertainment purposes instead of recognizing and highlighting truths.

References and Further Reading

Beesley, Lawrence. *The Loss of the SS Titanic: Its Story and Its Lessons*. Boston: Houghton Mifflin, 2000.

Carrothers, John C. "Lord of the Californian." *U. S. Naval Institute Proceedings* 94, 3, whole no. 781 (March 1968): 58–71.

Harrison, Leslie. *A Titanic Myth: The Californian Incident*, 2nd ed. Hanley Swan, Worcester, UK: S.P.A., 1992.

Heyer, Paul. *Titanic Legacy: Disaster as Media Event and Myth*. Westport, CT: Praeger, 1995.

Kuntz, Tom, ed. *The Titanic Disaster Hearings: The Official Transcripts of the 1912 Senate Investigations*. New York: Simon and Schuster Pocket Books, 1998.

McLean, Craig, and David L. Eno. "The Case for Captain Lord." *Naval History* 6, 1 (Spring 1992): 26–29.

Molony, Senan. *Titanic and the Mystery Ship*. Stroud, Gloucestershire, UK: Tempus, 2006.

Molony, Senan. *Titanic: Victims and Villains*. Stroud, Gloucestershire, UK: Tempus, 2008.

Padfield, Peter. *The Titanic and the Californian*. New York: John Day, 1966.

Williams, Thomas B. *Titanic and the Californian*. Edited and revised by Rob Kamps. Stroud, Gloucestershire, UK: Tempus, 2007.

3

The assassins of Archduke Ferdinand were funded by the Serbian government.

PRO   László Kocsis

CON   Steve Garrin

PRO

Napoleon unleashed nationalism on Europe and this led to the emergence of the Pan-Slavic movement among the Slavic people, which called for regaining lost Slavic supremacy in order to achieve unity of all Slavic peoples. Russia was the main ideological sponsor because it coincided with Slavic imperialist interests—unifying all Slavs under the tsardom of Russia and thus expanding the vast empire farther west, encompassing half of the "other half" of Europe. This is still reflected in the same three colors for the Czech Republic, Slovakia, Serbia, and Russia, adopted during the 1848–1849 Pan-Slavic congress held in Prague. The dream came true after World War II, when the Russian Empire, later the Soviet Union, seized control of half of Europe.

The Pan-Slavic movement considered the existence of Austria-Hungary as the main obstacle of uniting the territories inhabited entirely or partially by Slavs. The main focal point was the territorial unity of northern Slavic (Czech, Russian) and southern Slavic people, such as the Serbians, Croats, Bosnians, Bulgarians, and Macedonians. The Polish left this ideology as they rightfully saw Russia as the biggest threat to their national survival and not Austria-Hungary. Northern Pan-Slavism was led by the Czechs, and the southern stream was led by the Serbs, who have naturally concentrated their efforts on the Balkans, home of the southern Slavic people, not only where the significant non-Slavic population such as the Albanians, Hungarians, Vlachs (Romanians), and Greeks had been ruled from the medieval ages by the three great powers in the region but also in Austria, which embodied the Habsburg monarchy. In 1867 Austria became Austria-Hungary, the kingdom of Hungary until 1526, the Ottoman Empire until 1913, and the Republic of Venicel each had its sphere of influence. The southern Pan-Slavic movement advocated the independence and union of the Slavic peoples in the Austro-Hungarian Empire and the Ottoman Empire.

After the congress of Berlin of 1815, the Serbs gained autonomy from the decaying Ottoman Empire, but almost immediately after began seeking expansion to remove the other southern Slavs from under Serbian rule. The Pan-Serbian

49

movement gained momentum after 1878 when Serbia regained independence from Turkey, after losing the epic battle of Kosovo in 1389. Austria feared that Slav nationalists would endanger its empire, and this fear was justified as Pan-Slavism in the Balkans, contrary to Poland or the Baltics, often turned to Russia for support. Russia's victories over the Ottoman Empire made it possible for Serbia, Montenegro, Bulgaria, and Romania to become independent by the end of the 19th century.

Some Serbian intellectuals sought to unite all of the southern Balkan Slavs, whether Orthodox, Catholic, or Muslim, under their rule, transforming it into Pan-Serbism, an ideology that became prominent in the whole Serbian society and considered the supreme national goal. Pan-Serbism aimed at unifying all the Serbian people spread over the Balkans, and thus a Serbian-Nationalist movement was born. Pan-Serbism was fueled from anger over 500 years of Turkish rule, but it also became anti-Hapsburg, especially after Bosnia's annexation in 1908 by the Austro-Hungarian Empire. The Serbians considered Bosnia a part of their homeland and saw the annexation as a territorial conquest from Serbia, although in reality the Ottoman Empire left Bosnia under Austrian administration since 1878. The Pan-Serbism went even further, asserting the Serbian supremacy or preeminence over all Slavic people or nationalities of the Balkan Peninsula and asserting its right to conquer all the Slav lands just south of the Danube.

Kept at bay by Austria in the north, the Serbs directed their efforts toward the south on the Ottomans, and toward the east on the Bulgarian borders. After a stalemate with the Bulgars in 1885, they joined forces with the Bulgars and Greeks toward defeating the Ottoman Empire in first Balkan War in 1912. The Serb intelligentsia and diplomacy meanwhile were claiming that Macedonia had the essential Serbian identity, just as the opposite was claimed by the Bulgarians, which led to the second Balkan War in 1913, ending in a Serbian victory with the help of its allies: Greece, Romania, and Turkey, the latter seeking revenge over Bulgaria for previous territorial losses. Macedonia was divided between Serbia and Greece, Southern Trace was seized by Greece, and Turkey could hold Rumelia, its last foothold on the European continent. Greatly strengthened and with morale at its peak, in 1914, right after Pan-Serbism had been directed northwest toward Austria-Hungary, Bosnia-Herzegovina aimed to seize as much territory as possible from the monarchy, with the help of Russia. By 1914 Pašić had articulated a policy that Serbia would seek to incorporate all regions that contained Serb and other southern Slav populations, laying the foundation for what would become Yugoslavia. The idea was to unite lands that were or had historically been Serbian or where a significant minority of Serbs lived. The promoters of this idea cited a favorite slogan in the Balkans: "only unity can save the nation, in this case the Serbs."

In Austria-Hungary, southern Slavs were scattered among several entities: Slovenes were living in the Krajina region of Austria, and Croats were inhabiting Istria and had an autonomous province called the kingdom of Croatia-Slavonia, in

personal union with the kingdom of Hungary, since 1100. In other parts like Bosnia-Herzegovina they lived mixed with Serbs and Bosnian Muslims, being under direct Austrian control. Croats were dominating the Austrian autonomous kingdom of Dalmatia. In Austria-Hungary, the idea of Austroslavism was attractive, which sought to unite Austrians with the Slavic peoples outside of the state of Serbia.

The greater Serbia ideal was formally launched by Serbian minister Ilija Garašanin's work, called the Načertanije, created in 1844, which aimed at uniting the Serbian people from across Eastern Europe. This plan, hypothetical in the extreme, was taken quite literally by later extremists, such as the terrorists in the organization that came to be known the Black Hand.

Nineteenth-century Serbian linguist Vuk Karadžić claimed that all southern Slavs that speak the štokavian dialect are Serbs who speak the Serbian language, emphasizing what united southern Slavs rather than the religious differences that had divided them. Therefore, it has often been suggested that the Muslims of Bosnia are the descendants of Serbs who converted from Orthodox Christianity to Islam under the rule of the Ottoman Empire; a similar claim was made by the Croats but with Catholic, therefore Croat, origin.

Ironically the Habsburg Empire at the beginning supported certain unification efforts among the Slavs, financing the formation of the Matica srpska (Serbian Mother) in 1826 and supporting the Serbian Orthodox Church, but it soon came to oppose Pan-Slavismas, a detrimental factor to its own unity. The government in Vienna became suspicious when these institutions turned into political propaganda machinery aimed at secession and Serbian expansion into Austrian territory.

As Pan-Serbism was the creator of Yugoslavia, it also had the biggest role in its final dissolution. The memorandum of the Serbian Academy of Sciences and Arts from 1986 set in motion the Pan-Serbian movement of the late 1980s by claiming alleged systematic discrimination against Serbs and Serbia, culminating in the Serbs of Kosovo becoming the victims of genocide, Yugoslavia being under Croatian-Slovene hegemony, and the Serbs being oppressed in Yugoslavia as a nation. The oppression was especially brutal in the Serbian autonomous province of Kosovo-Metohija and in Croatia, where their status is "the worst ever as far as recorded history goes." The memorandum claimed that Serbia was being economically exploited, with much of her wealth being drained off and redistributed to the other republics. The existing borders between the Yugoslav republics are claimed to be drawn arbitrarily by dominant Croatian and Slovene communists, motivated supposedly by anti-Serbian feelings. The Croatians and Slovenes accused Serbia for apathy toward non-Serbian interests, for example, in the case of the kingdom of Serbs, Croats, and Slovenes ceding Istria to Italy, a move that angered local Croatian citizens and left Slovenia with no access to the sea after World War I.

The Serbian society claimed the necessity of the breakup of Yugoslavia on greater Serbian lines was to readjust Yugoslavia and to redress the alleged bias against Serbia. The greater Serbia of the early 1990s included all or most of Croatia,

all of Bosnia, Montenegro, Macedonia, and also naturally called for the abolition of the autonomy of Serbian provinces of Kosovo and Vojvodina, rejecting the existing boundaries as the artificial creation of Tito's partisans. This led to the rise of Slobodan Milošević, who in turn supported the extremely nationalistic groups, insisting on the demand for "all Serbs in one state." The means of achieving it was the widespread use of terror and ethnic cleansing to ensure that ethnically homogenous territories could be established, without opposition from potentially disloyal minority groups. The ideal didn't take into account, however, that the borders of present Serbia are still questionable, on similar grounds, since the vast majority of Albanians, Bosnians, or Hungarians who are citizens of Serbia living mainly in Kosovo, Sandhzak, and Vojvodina regions want, naturally, to live in their respective national states. When Kosovo broke away, Serbia used this time as a softer ethnic cleansing in Vojvodina to hold this territory, making psychological pressure unbearable for the Hungarian minority and forcing them to emigrate from their ancestral land and bringing hundreds of thousands of Serbian refugees from Croatia and Bosnia.

The original idea of Europe-wide Pan-Slavism has vanished after the 1990s, failing to establish and harmonize Czechoslovakia and Yugoslavia, but it still has an appeal in Russia, Serbia, and Slovakia. As a result of the wars accompanying the breakup of Yugoslavia, hundreds of thousands of southern Slavs became victims of genocide, committed by their fellow Slav "brothers," mostly in Bosnia, with the great majority of the perpetrators being Serbians. The accused or convicted leaders, most of them Serbs, were tried by the International Criminal Tribunal for the former Yugoslavia (ICFY) for war crimes and crimes against humanity, proving that the chauvinistic irredentist drive that led to the assassination of Franz Ferdinand and the two world wars can cause only sufferings and offers no solutions even one century after.

International Developments in the Power Relations, before the Assassination

The French–Russian secret alliance enlarged later with Britain and the entente cordiale in 1907 convinced the Emperor Franz Joseph that only Germany could offer suitable support for Austria. The 1908 annexation of Bosnia-Herzegovina from the Ottoman Empire sparked an international crisis, Russia having felt itself betrayed by Austria, who in turn promised free access for the Russian fleet through the Dardanelles, even in cases of war. In exchange for the annexation, however, Austria-Hungary gave up the Sandzhak in 1909, and Britain finally agreed, also fearing Mediterranean rule threatened by the Russian presence. Germany put pressure on Russia and the Ottomans to accept annexation, and the sultan received serious monetary compensation for it from Austria.

The annexation was bitterly opposed by Serbia, who targeted Bosnia as the future most probable area of territorial expansion. The annexation was also a

priority opposed by the Hungarian politicians of the dualist monarchy, considering it an unnecessary provocation for Slavic nationalism. It has been proven that both the Serb and the Croat nationalists were working on the creation of a greater Serbia and greater Croatia, respectively, and both wanted to incorporate Bosnia fully into their empire. Although outraged by the annexation, which also profited Serbia, it took possession of the Sandzhak together with Montenegro after the Austro-Hungarian withdrawal from that region as an exchange for the annexation, shortly after Serbia apparently took a conciliatory position. On March 31, 1909, it committed itself without sincerity to respect the territorial integrity and sovereignty of Austria-Hungary, to drop its attitude of protest with respect to the newly annexed provinces, and to maintain good neighborly relations with the monarchy. The acceptance was carried out mainly for diversion purposes in order to distract attention about its real goals—that of the conquest of Bosnia and many more Austro-Hungarian territories. The annexation further weakened the Ottoman's prestige as the day following it Bulgaria declared its full independence, so the potential enemy lines had been drawn fairly close to Istanbul.

In 1911 the Balkan alliance was created under Russian hegemony, which principally was directed toward the liberation of the people under Ottoman rule but also against Austria-Hungary. The Romania participation in the second Balkan War was signaling Romanian expansionist goals into Transylvania at the expense of Hungary. In June 1913 Serbia and Romania concluded a secret agreement in Bucharest for the annexation of large Hungarian territories. Hungary's Prime Minister Tisza István tried to forge a Hungarian–Romanian understanding, but this hurt Russian and French interests, leading to the assassination of the Hungarian Greek Catholic bishop in Debrecen. The Russian diplomacy promised Transylvania for Romania, while Vojvodina and any other Hungarian or Austrian territory the Serbs would be able to conquer for Serbia.

Prior to the assassination in 1914, Serbia had begun preparations for war. In 1911 it set up three listening posts on the Austro-Hungarian border and drew up military cartography of the border region. The Serbian border guard officials also had the unofficial tasks of organizing revolutionary groups in Bosnia, such as the Narodna Odbrana, and forming contacts with Bosnian agitators in order to both enroll volunteers and to raise an insurrection in the event of war with Austria-Hungary. In the following years, a series of assassination attempts were made against Austro-Hungarian officials in Croatia and Bosnia-Herzegovina, culminating in the Sarajevo events of June 1914.

The International Political Objective of the Assassination—Irredentism

Austria-Hungary received the mandate to occupy and administer the Ottoman provinces of Bosnia and Herzegovina, under the Treaty of Berlin in 1878,

which also recognized Serbian independence, while the Ottoman Empire retained official sovereignty. Under the same treaty, Serbia was at last recognized by the great powers as a fully sovereign state, as the kingdom of Serbia. Initially Serbia was content to live within its smaller territory, which encompassed around two-thirds of the ethnic Serbian population. This annexation was viewed with envy by Serbia who hoped to seize Bosnia, but having no military power, it organized subversive organizations like the Narodna Odbrana. This semisecret society called the Defense of the People was part of the institutional network of the Pan-Slavic nationalist movement dedicated to achieve its nationalistic goals by means of assassination.

By 1911 the organization has split tactically. The political leaders in the Serbian Radical Party preferred a more passive approach for the time being with Austria, concentrating on preparing Serbia for the future struggle, but some nihilist army officer members grew impatient, and the more zealous conspiratorially-minded ones founded a new secret society in the kingdom of Serbia in May 1911, the Black Hand (Crna Ruka). The purpose of the nihilist group was to recruit and train partisans for a possible war between Serbia and Austria and eventually free Bosnia from Austria. They organized spies and saboteurs to operate within the empire's provinces including Bosnia and Herzegovina. The society officially known as Unification or Death encompassed different extremely nationalistic ideologies, being similar to the Russian People's Will and the Chinese Assassination Corps, which considered assassination the most suitable means for achieving political goals.

Members of this death squad comprised Serbian army officers who were dissatisfied with the 45-year-long rule of the Obrenović dynasty and especially the authoritarian and outrageous rule of King Alexander I and his wife, Queen Draga, and they murdered him in cold blood in 1903. Alexander's downfall was his policy of strict neutrality in the 1897 Turkish-Greek war, when he should have supported the Greeks, according to the Serbian army's opinion. The 26-year-old king was considered a fool garbed by a wicked seductress, Draga, who was 15 years older than Alexander and was the widow of a Czech engineer and a lady-in-waiting to Alexander's mother, Queen Natalija. The queen also bitterly opposed the marriage, but she was sent into exile by her son, the former king, Milos, who resigned earlier and also quit his office of the commander of the army. Alexander I's many arbitrary and unpopular acts were blamed on Draga's influence, and the prospects that the widow's brother would be nominated heir to the throne in case the marriage should be proven fruitless inspired the Black Hand leader Colonel Dragutin Dimitrijević, nicknamed "Apis," to become the perpetrator of the June 1903 coup, which resulted in the butchering of the royal couple in their palace at Belgrade. The assassination had brought the rival King Petar Karađorđević to the Serbian throne, which was considered much more nationalistic and therefore suitable for the purposes of the army.

In 1909 Austria began to apply pressure to the Serbian authorities to deal with the anti-Austrian elements in their society. The Serbians, with no one else to turn to, had no choice but to comply, so the movement temporarily took a step away from direct action toward education and winning the hearts and minds of the people. In the first Balkan War Russia backed Serbia, the mainly Slavic nations of the Balkan, including Romania, in the name of Pan-Slavism but also Greece against the Turks, while the Western powers included Great Britain, Germany, the Austro-Hungarian monarchy, and Italy that were interested in keeping Turkish territorial integrity. The 1912 London agreement divided the contested territories inhabited by predominantly Muslim or other Slavic people like the *vilayet* of Sandzhak, Kosovo, which included Skopje, Kumanovo, and Tetovo, now part of the Republic of Macedonia, as well the whole of Macedonia between the victors. Emboldened by successes in the Balkan Wars of 1912–1913, during which it seized Kosovo and northern Macedonia, Serbia in 1914 wanted to break Austria-Hungary's southern Slav provinces off and whatever it could conquer from the rest so they could be combined into a greater Serbia, which eventually would be called Yugoslavia.

Why Was Franz Ferdinand and Assassination Chosen?

Given the timing—the year of 1914—and the series of Serbian successes even in the years preceding it, it was clear that Serbian politicians were determined to achieve their political goals at any cost, and now, without delay or hesitation. The archduke and heir to the throne Franz Ferdinand, through his Czech fiancée, was becoming increasingly a supporter of a trialist monarchy, including the Slavic people of the monarchy under Czech leadership or together with the resurrection of a greater Croatia, including Dalmatia and Bosnia. The realization of such a trialist confederation, by combining the Slavic lands within the Austro-Hungarian Empire into a third crown, would have made greater Serbia impossible to achieve, with its envisaged borders, so Franz Ferdinand became the only logical target.

This was becoming even more evident in the light of the series of tragedies suffered by the family of the Habsburg Emperor Franz Joseph. The queen, Elisabeth of Bavaria, was killed by an Italian anarchist a few years before their only son, Rudolf, was found dead with his Czech girlfriend Vetshera in Meyerling after a mysterious, supposedly suicidal act. The emperor himself was old, serving already for 66 years, at an age of 80 in 1914, so he was expected to die soon, which did happen two years later in 1916. The heir to the throne Franz Ferdinand planned to create a Slavic kingdom within the monarchy as a bulwark against Serb irredentism, and therefore he was perceived as a threat by the same stream of irredentists. Gavrilo Princip later stated to the Austrian court in Sarajevo that prevented Franz Ferdinand's planned reforms was one of his motivations.

Technically, the governor of Bosnia Oskar Potiorek was the first target, but the commissioned murderer, Bosnian Muslim Mehmedbašić, failed to prepare the assault from fear and probably remorse. On March 26, 1914, the organizer of the plot on the field Danilo Ilić informed Mehmedbašić that Belgrade had scrapped the mission to kill Potiorek and its plans now were to murder Franz Ferdinand, so he should wait for new orders. The assassination was ordered by the head of the Intelligence Department of the Serbian Army, Colonel Dimitrijević (Apis) acting in his official position, as he confessed to the Serbian Court set up in Saloniki in 1917. Major Vojislav Tankosić took part in the plotting. He was a prominent member of the Black Hand, who was in charge of guerrilla training in the Serbian army by that time. He reached an agreement to transport arms from Belgrade to Sarajevo with fellow Bosnian and former guerrilla fighter and known arms smuggler Milan Ciganović, who promised help in recruiting volunteer murderers. At first, Tankosić provided the assassins one pistol to practice with, the rest of the weapons being delivered later, on May 26. Franz Joseph, the old emperor, was ill, and rumors spread that the heir apparent would not go to Bosnia, therefore the operation was temporarily halted, but after the emperor recovered the operation started again.

Franz Ferdinand made an unfortunate choice by visiting Sarajevo. He was somehow forced to do so, but because of his "unsuitable" choice of consort, he could not take his wife with him on official visits, excepting military missions. That's why he had chosen Sarajevo as the place to inspect the troops, on the 14th anniversary of their morganatic marriage. His love for Sophie made him blind to the dangers, and he decided to visit the turbulent, recently annexed province in an open car, unprotected, amid the warnings of possible assassination attempts. The timing, June 28, was also very unfortunate since it was also the date of Serbian national mourning, called Vidovdan or Saint Vit's Day, when the first battle of Kosovo was lost in 1389 and the medieval Serbian statehood ceased to exist. The anniversary evidently called for patriotic acts eventually resembling those of Miloš Obilić, the victorious sultan who had been assassinated on that same day, as a deserter walked into the sultan's tent, leaving him unprepared.

Another reason for these actions was the Serbian tradition of assassination of royals or important politicians, a tradition that dates back as early as 1389 and continues to the present day. Among a series of failed attempts against Austrian officials, in 1903 the Black Hand, in fact led by Colonel Apis, stormed the Serbian Royal Palace and brutally assassinated King Alexander I and Queen Draga, shot 30 times then eviscerated and mutilated their body parts and threw them out the window. The royal guard general Petrović was also killed. Tankosić organized the subsequent murders of Queen Draga's brothers. In the five years prior to 1914, lone assassins—mostly ethnic Serb citizens of Austria-Hungary—made a series of unsuccessful assassination attempts in Croatia and Bosnia-Herzegovina against Austro-Hungarian officials. The most famous of these was Bogdan Žerajić's attempt on June

15, 1910, which aimed to kill the iron-fisted governor of Bosnia-Herzegovina, General Marijan Verešanin. Žerajić was a 22-year-old orthodox Serb from Herzegovina who made frequent trips to Belgrade. Just 12 days before the attempt, Žerajić had made an aborted attempt on the life of Emperor Franz Joseph. The failed assassination attempt and the subsequent suicide of Žerajić gave inspiration to future Serbian assassins, including Princip and his accomplice Čabrinović. Princip stated that "Žerajić was my first model. When I was seventeen, I spent whole nights at his grave, reflecting on our wretched condition and thinking of him. It was that place where I made up my mind to perpetrate an outrage, sooner or later."

Colonel Apis, as member of the general staff of the Serbian army, decided to become a specialist in terrorism. In 1911 he organized an attempt to assassinate Emperor Franz Josef. When this failed, he turned his attention to the heir to the throne, Archduke Franz Ferdinand. When he heard that Franz Ferdinand was planning to visit Sarajevo, he sent three members of the Black Hand group, Princip, Nedeljko Čabrinović, and Trifko Grabež from Serbia, to assassinate him. The assassins' motives were consistent with the movement that later became known as Young Bosnia, encompassing individuals, groups, and organizations like the Narodna Odbrana, and the Black Hand, who favored the violent destruction of Austria-Hungary to make way for the unification of the southern Slavs with Serbia.

Apis had personal expertise in assassinations; the Serbian Parliament praised him as "the savior of the fatherland" after his 1903 rampage. Similar recognition was granted by many to the assassins of Serbian Prime Minister Zoran Gingic in 2003, considered a traitor, who had handed over national hero and war criminal Slobodan Milosevic to the ICTY in the Hague, for trial on charges of genocide and war crimes.

Before that, Milosevic successfully assassinated his opponents with his secret police, as in 2000 his mentor and close personal friend the former Serbian President Ivan Stambolić, also a greater Serbian nationalist, made his way to the Yugoslav presidency. The murder of the ex-friend was carried out by eight special operations unit officers. In spite of, or as praise for, the assassinations carried out "in the name of the Serbian nation" by Milosevic targeting less radical or less nationalist politicians, he was considered to be the savior by many who had eliminated those considered too soft on minorities and those who opposed his policies in any way. Nowadays ICTY-imprisoned Serbian Radical Party leader Vojislav Šešelj compared Gingic's assassin Zvezdan Jovanovic, a former lieutenant colonel in the police and member of the feared paramilitary unit, the Red Berets, with Archduke Franz Ferdinand's assassin, saying "he would enjoy the same fame and glory that Princip has in Serbian history."

Another recent example is the 25-year-old Serbian immigrant Mijailo Mijailović, born in Sweden who became the assassin of Swedish foreign minister Anna Lindh in 2003, apparently because he hated the minister for backing the NATO air strikes

against Serbia during the 1999 Kosovo War, which he considered an "injustice in the world."

Serbian Prime Minister Pašić in 1914 supported the main objectives of the group and the plot to kill the heir, but allegedly didn't want to go that far immediately as he feared it would lead to a war with Austria-Hungary. Therefore, he allegedly gave instructions at the last moment for the arrest of Princip, Čabrinović, and Grabež if they attempted to leave the country. However, his orders were not implemented and the three men arrived in Bosnia-Herzegovina where they joined forces with the fellow conspirators. This story seems untrue, since it presumes that in a centralized state the prime minister cannot control the border guards. And if he had opposed the plot, why wouldn't he have waited until Serbia's defeat in 1917 and tried the masterminds such as Apis only after the defeat of the country?

Serbia didn't care much for the foreign public opinion, being aware of the backing of Russia. This was apparent in 1995 in Bosnia, where the Serbs took hostage 370 peacekeeper troops and used them as live shields, leading to the first real-combat air raids of NATO. History has repeated itself in the respect that the court of the Hague has established in 1999 and given the utmost responsibility of trying the Yugoslav Serbian ex-president Milosevic in the ethnic cleansings and crimes against humanity committed in Bosnia, Kosovo, and Croatia; but only four years prior, in 1995, the same Milosevic was representing the Serbs in the peace talks taking place in Dayton, Ohio, and was praised by the West for stopping Bosnian Serb leaders Ratko Mladic and Radovan Karadžić. This inevitably draws the conclusion that the latter were acting in close cooperation and with the approval of Belgrade, otherwise why would Milosevic have represented them at the peace talks in Dayton? Similarly, the highest level Serbian official involvement was happening in 1914, in Bosnia as well, thus confirming the direct responsibility of Serbia.

The Serbian Government's Involvement in the Assassination and Its Russian Backing

Franz Ferdinand and his wife, Sophie, were shot dead while traveling in Sarajevo, the capital of Bosnia, by Princip, one of a group of six assassins coordinated by Danilo Ilić, after a careful planning of the murder was conceived at Belgrade. The murderers received the shotguns and the hand grenades from active Serbian officers and officials. The dispatch of the criminals and of their weapons to Sarajevo was arranged and effected under the control of the Serbian frontier authorities. In 1917 Apis, in his written statement to the Serbian Court, stated that he had ordered the assassination of Franz Ferdinand. In the assassination process, he used not only his official power over elements of the Serbian military, but also some Black Hand members. Leaders of the Black Hand in turn had penetrated the Narodna Odbrana and used the organization to infiltrate the arms and assassins into Sarajevo.

Archduke Franz Ferdinand and his wife in Sarajevo on June 28, 1914, moments before they were killed by an assassin working with the Black Hand terrorist group. (The Illustrated London News Picture Library)

The Guns that Started World War I

Four handguns were recovered from the members of the Black Hand, including the one that was used by Gavrilo Princip to shoot Archduke Franz Ferdinand. They were all exhibited at the subsequent trial, and after that, they were given to Father Anton Puntigam, a Jesuit who had planned a memorial to Franz Ferdinand, possibly through an orphanage at Sarajevo. Obviously after World War I, this became impossible, and the Jesuits kept the guns at their headquarters in Vienna.

In 2004 the guns were handed over to the Heeresgeschichtliches, the museum of military history, in Vienna, but there were soon claims that one of the guns was not a genuine one, and that another had been substituted. Since then various collectors have claimed that they own the actual gun Princip used.

The car in which Franz Ferdinand had been traveling when he was shot also survived. It was owned by Count Harrach and lent to Franz Ferdinand, but after the shooting the Harrach family did not want the car back, and it ended up in the Heeresgeschichtliches, the museum of military history, in Vienna. During the 1990s the Harrach family tried to get the car back through a court challenge, but this failed as it had been handed over to Emperor Franz Josef, who had, in turn, given it to the museum.

The Austro-Hungarian investigation of the assassination rounded up all but one of the assassins and also much of the underground channel that had been used to transport the perpetrators and their weapons from Serbia to Sarajevo. Under interrogation, the assassins fingered members of the Serbian military. Three of the assassins, who came from Belgrade, Serbia, confessed to the involvement of active duty members of the Serbian army and Frontiers Service. The events of June 28 have proved the existence of a subversive movement in Serbia, whose object was to separate certain portions of territory from the Austro-Hungarian monarchy. This movement, which came under the purview of the Serbian government, subsequently found expression outside the territory of the kingdom in acts of terrorism, through a number of assassination attempts and murders. Far from fulfilling the formal obligations contained in its declaration of 1909, the royal Serbian government has done nothing to suppress this movement, rather it has been guilty of tolerating the criminal irredentist activities against the monarchy of the various unions and associations, allowing the glorification of the assassins.

The Russian backing of Serbia's actions came to light when Apis revealed that the Russian military attaché had agreed to protect Serbia from Austria-Hungary and to fund an assassination attempt against Austro-Hungarian targets. Although the Russians tried to deny their involvement, the assistant military attaché, who took over while the head attaché was on vacation, admitted their involvement with Apis.

There is evidence that Russia was at least aware of the plot prior to June 14, since on June 1, 1914, in Emperor Nicholas's meeting with King Charles I of Romania at Constanza, Russian Foreign Minister Sazonov mentioned that he was convinced that if the archduke (Franz Ferdinand) was out of the way, the peace of Europe would not be endangered. After its publication, the entente apologists argued that "out of the way" might not have necessarily meant assassinated.

What was apparent is that the Serbian government carried out subversive propaganda and has tolerated every manifestation of hatred against the monarchy and contempt for its institutions. Serbian officials, whether at home or abroad, had made hostile statements toward the monarchy prior to the assassination. So the German and Austro-Hungarian claims that the assassins were funded by the Serbian government were later fully proved. This proof was strengthened by the events following the assassination.

Within the following two days, Austria-Hungary and Germany advised Serbia that it should open an investigation, but secretary general to the Serbian foreign minister Sava Grujić, in his official position, replied falsely on June 30: "Nothing had been done so far and the matter did not concern the Serbian Government." He acted so with evident Russian backing as the declaration was made after receiving assurances from Russian Foreign Minister Aleksander Izvolsky and Tsar Nicholas II that Russia would solve the Serbian question by bringing about the downfall of Austria-Hungary, once it had rebuilt its army.

Izvolsky was particularly upset with Austrian foreign minister Alois Aehrental, whom he perceived to have tricked him into the secret talks held at Buchlau (Buchlov) before the 1908 annexation. Germany tried to put a halt to the escalating verbal conflict by asking Russia to intercede with Serbia, but its request was ignored. The Russian backing for this provocation was evident when Austria-Hungary and Germany asked Serbia directly and through her ally Russia to open an investigation into the Sarajevo outrage; Russia and Serbia flatly rejected these requests.

Following the assassinations, Serbian ambassadors to France, Milenko Vesnić, and to Russia, Spalaiković put out statements claiming that Serbia had warned Austria-Hungary of the impending assassination. Serbia soon thereafter denied making warnings and denied knowledge of the plot by its Prime Minister Pašić. The truth, however, is that Pašić was fully aware of the assassination, as Serbian education minister Ljuba Jovanović wrote in his book *Krv Sloventsva*. In late May or early June 1914, Pašić reviewed the impending assassination plot with members of his cabinet. On June 18 a generally worded telegram ordered Serbia's ambassador to Vienna Jovanović to warn Austria-Hungary that Serbia had reason to believe there was a conspiracy to assassinate Franz Ferdinand in Bosnia. On June 21, 1914, Jovanović met with Austro-Hungarian finance minister Bilinski, stressing in general terms the risks the visit may bear justified by the inflamed public opinion in Bosnia and Serbia, stating that "Some serious personal misadventure might befall him. His [FF] journey might give rise to incidents and demonstrations that Serbia would deprecate, but that would have fatal repercussions on Austro-Serbian relations," but the finance minister took no action based on these vague and misleading remarks.

Jovanović went public in 1924, stating that the warnings were made on his own initiative, but an ambassador never warns another country from his own initiative about such a serious question, because he is immediately dismissed, and the person is probably tried for high treason. Moreover, normally an ambassador wouldn't say that "Among the Serb youths (in the army) there may be one who will put a ball-cartridge in his rifle or revolver in place of a blank cartridge and he may fire it, the bullet might strike the man giving provocation [Franz Ferdinand]." Jovanović's account changed back and forth over the years, while Bilinski did not speak openly on the subject, but his press department chief confirmed that a meeting had taken place and included a vague warning.

The July Ultimatum and the Serbian Reaction

The Serbian involvement was strengthened by the country's further attitude toward the consequences of the assassination. On July 23, Austria-Hungary delivered a demarche to Serbia, known also as the July Ultimatum. The monarchy waited with its dispatch until 6:00 P.M., when the Franco-Russian summit would

have ended. The demarche contained 10 enumerated and some additional demands in the preamble and was completed by an annex comprising sufficient criminal evidence, describing the findings of the criminal investigation into the Sarajevo events in which assassins sent from Serbia killed the royal couple and wounded 20 people. The ultimatum called for the destruction of the anti-Austrian terrorist and propaganda network in Serbia, based on Serbia's 1909 declaration to pursue a good neighbor policy.

The ultimatum did not include any explicit threat of war, only a threat to recall Austria-Hungary's ambassador if Serbia did not accept the demands within 48 hours. The note demanded that the Serbian government give an official, precisely worded assurance that admitted the existence of anti-Austro-Hungarian propaganda and the participation of Serbian officers and functionaries in it, and that this proclamation should be published as an order of the day by the king to the army and also on the first page of the July 26 issue of the "Official Bulletin of the Army." Serbia was called to condemn and "suppress with all the means at its command this criminal and terrorist propaganda and express the most sincere regrets for such involvement of Serbian officers and officials and its dreadful consequences," which have endangered the friendly and neighborly relations, to the cultivation of which the Royal government had most solemnly pledged itself by its declarations of March 31, 1909. Serbia should warn future perpetrators that the Serbian government would crack down against such activities with utmost rigor.

The note accused Serbia of orchestrating anti-Austro-Hungarian propaganda and terrorist actions, assertions proven by evidence, while Serbia rejected many of them, asking for further evidence of his involvement. In the end, Serbia grudgingly and seemingly accepted most of the demands but didn't want to implement many of the measures promised or as they were formulated in the demands, using ambiguities and reserving its right to declare anything after the transmission of the answer. After receiving a telegram of support from Russia, Serbia mobilized its army and responded to the letter by accepting points 8 and 10 entirely and partially accepting, finessing, ingenuously answering, or politely rejecting the other demands 1–7 and 9. It rejected point 6, asserting that it would have compromised its sovereignty.

The 10 listed demands addressed to Serbia can be grouped into the following five categories. The first group of demands was that Serbia should suppress publications that incite hatred and contempt, eliminating from public instruction in Serbia those teachers and methods of instruction that served to foment propaganda against Austria-Hungary. The Serbian reply stated that currently it was impossible to control its press given Articles 12 and 22 of its Constitution, but it would pass a bill against inciting hatred. Without specifying any deadline, it was unclear whether this bill would ever be passed, and if so, if it would ever be implemented or enforced. In fact, major newspapers were always under some government control, so this assertion is in fact untrue.

The second group of requests was to dissolve the terrorist organizations like the Narodna Odbrana active in Bosnia, confiscate their means of propaganda, and prevent these societies from reforming under new names. Serbia should remove from the military service and the administration in general all officers guilty of such propaganda, some names being expressly provided by Austria-Hungary. Such organizations were the Black Hand and the White Hand (Bela Ruka) formed in 1912 by Colonel Petar Živković, involved also in many coups like that of March 27, 1941. The main one, the Narodna Odbrana, had its central committee located in Belgrade, had organized paramilitary forces, and expressed a need for the protection of the nation. Its 1911 pamphlet asserted that hatred of Austria-Hungary was not the aim of the group, but instead it was a natural consequence of the desire for independence and unity of Serbs within a single state and presented six main points among, which the second was "registration and enlistment of volunteers"; the third, "formation of volunteer units and their preparation for armed action"; the fourth, "collection of voluntary contributions, including money and other things necessary for the realization of its task"; the fifth, "organization, equipment, and training of special revolutionary bands for special and independent military action," which now would be considered the bylaw of any terrorist organization. The Austrian demarche also called for the change of political rhetoric after June 28 and for the termination of the hostile propaganda, but as propaganda was not yet a crime in Serbian law, it remained unclear how many officers should or would be removed.

The third group of demands focused on the acceptance of the collaboration of organs of the Austro-Hungarian government carried out in Serbia for the suppression of the subversive movement; and under point 6, Serbia should begin a judicial inquiry against those perpetrators of the plot who are on Serbian territory, with organs delegated by the Austro-Hungarian government participating in the investigation. Point 6 was later rejected by Serbia, calling it a violation of international law. This in fact is not the case since police cooperation in transboundary crime investigations is commonplace; moreover the drafts of the Serbian reply had originally contained the acceptance of this point with reservations, but after receiving Russian assurance that Serbia would be supported even if it did not fully comply with the demands, their reply became tough.

The Serbian reply stated that it would cooperate with the Austro-Hungarian police so long as this was consistent with international law, criminal procedure, and good neighborliness, but in its characteristic ambiguity avoided any clear answer in order to gain time. Serbia said that it would open an inquiry, but an involvement of Austro-Hungarian officials in this inquiry would be unconstitutional and a violation of the national criminal law. Serbia primarily expressed surprise over the "presumably supposed" involvement of its citizens in the Sarajevo assassination, which was clearly a lie since the plot was organized on high official level by the intelligence of the military. Prime Minister Pašić knew and

approved it, so it was unclear why he had taken no action. Another response was that Ciganović was an Austro-Hungarian citizen so he was not under its jurisdiction. The Serbian reply stated falsely that Serbia had no proof that the Narodna Odbrana and similar societies carry out paramilitary activities, but if it could be proven it would nevertheless disband them, thus maintaining a free hand in every aspect. The Serbian reply also failed to address aspects such as confiscating the means of propaganda of these organizations and preventing them from reforming under new names, suggesting that really nothing would have been done.

The fourth group of demands asked Serbia to arrest Tankosić and Ciganović immediately, as they had been identified as the assassins. Serbia replied that it had arrested Tankosić but stated that Ciganović could not be found. Austria-Hungary pointed out that the reason he could not be found was because the Belgrade chief of police had told him to go into hiding. Such practices are well known even today, when Belgrade conceals the biggest Bosnian war criminals of 1992–1995 by not revealing the hiding places of Mladic and Karadžić, the most wanted by the ICFY, who are still at-large. Only very recently Belgrade seeking European Union accession gave in partially to the demands, "finding" former republikas Sprsak President Karadžić, but not general Mladic who may leak sensible political and military information to the judges.

In fact Serbia released Tankosić promptly after arresting him and he returned to his unit. Another perpetrator, Mehmedbasic, escaped to Montenegro where he was arrested. Austria-Hungary asserted its right to extradite him, but Montenegrin authorities instead allowed him to "escape" to Serbia where he joined Tankosić's unit. Another mind of the plot, masterspy Rade Malobabić, the Serbian military intelligence's top agent against Austria-Hungary, was arrested on his return from Austria-Hungary's Bosnia after the assassination, but was later released and given a commission to run an army supply store. He was, however, executed after the Saloniki trial, but on charges of perceived high treason, not because of his involvement in the Sarajevo plot.

In the last group of demands Serbia formally agreed to act in order to prevent the illicit trafficking in arms and explosives over the Austro-Hungarian border but didn't address the request for the dismissal and severe punishment of the Serbian borders service officers who assisted the assassins in smuggling the weapons to Sarajevo.

This noncompliance with the demarche was subsequently considered to be a legitimate *casus belli* against Serbia. The rejection of the ultimatum by Serbia is an indirect proof of Serbia's involvement as it pursued this outcome of the plot. Austria considered with good reason that the facts had been proven and if this was not admitted by Serbia, the defendant had to prove the contrary. As it could be foreseen, Austria-Hungary recalled its ambassador.

In fact Serbia admitted tacitly its guilt and bad intentions by ordering general mobilization on July 25, before sending any response that was sent only

just before the deadline expired. Serbian reservists were transported on tramp steamers on the Danube, apparently accidentally, and made a provocation by crossing into the Austro-Hungarian side of the river at Temes-Kubin, so the empire's soldiers fired into the air to warn them off. Given the extremely tense political situation, with the rejection of the demarche, this incident was blown out of proportion, and Austria-Hungary declared war and mobilized its army on July 28, 1914. Under the Secret Treaty of 1892, Russia and France were also obliged to mobilize armies if any of the Triple Alliance of Germany, Austria-Hungary, and Italy had mobilized.

That is exactly what Serbia wanted to provoke, backed by Russia and driven by Pan-Slavism and the slogan of liberating the oppressed Slavic people of the Austrian monarchy, it in fact was looking to be the main power in the Balkans and acquire the rest of Poland (Galicia) and use the Czechs as pawns in the German region stretching from the Baltic to the Danube and Alps. Russia mobilized immediately, evoking the treaty with France, and soon all the great powers of Europe were at war, except Italy. Italy was invoking a clause in the Triple Alliance Treaty that only bound it to enter in case of aggression against one of the treaty members, and so it remained neutral, for the time being, but it changed sides in 1915 to fulfill its irredentist goals.

In 1916 Serbia and Romania were defeated and in 1917 Russia too, asking for armistice, and Austria-Hungary thought the war had been won. Romania was secretly both an ally of Austria-Hungary until 1916 as well as of Russia, the opposing forces having no knowledge of this double agent behavior. Russian ambitions were clearly shown by its promise of upper Hungary (today's Slovakia) and northeastern Hungary-Karpatalja or Zakarpatska oblast (today in Ukraine) to Czechia, which was to be ruled by a Russian crown prince. After the victory of 1916, Austria wanted to annex Romania, but the Hungarians vetoed this. Hungary supported the unification of Dalmatia with Croatia. After being in the East, Poland, Ukraine, and the Baltics becoming independent, and the unjustified intervention of the United States on the side of the entente decided the war otherwise, rescuing the Soviets in November 1918. This intervention led to a false sense of victory among the British and revenge of the French who, among others, made the Pan-Slavic dream come true, creating Yugoslavia, Czechoslovakia, and greater Romania, which broke up shortly and every time thereafter. These newly created French satellite states "had the task" of eradicating German and Austro-Hungarian influence in the region, weakening the latter so much that they could never recover. France forbade the continuation of the monarchy, forcing Emperor Karl to abdicate in 1918, when he tried to get crowned as apostolic king of Hungary.

The concept of "greater Serbia" was put in practice during the early 1920s, under the premiership of Pašić, who, using tactics of police intimidation and vote rigging, diminished the role of the oppositions, mainly those loyal to his

Croatian rival, Stjepan Radić. He created a centralized power in the hands of the Serbs in general and Serbian politicians in particular, a policy continued unabated by other Serbian politicians, most recently by ex-Yugoslav president Slobodan Milosevic, father of failed efforts to make a unitary Serbian state during the 1990s. This experiment has cost hundreds of thousands of lives in one of the world's worst genocides since World War II, adding to the many million victims of the useless world wars in Europe.

Crime as Revealed by the Sarajevo and Saloniki Trials and the Capture of Serbian Archives

The Sarajevo investigation of the crime revealed that the plan of murdering Franz Ferdinand was directly concocted in Belgrade by Gavrilo Princip, Nedeljko Čabrinović, a certain Milan Ciganović, and Trifko Grabež with the assistance of Major Tankosić. Members of the Black Hand group under interrogation by the Austrian authorities claimed that three men from Serbia—Dimitrijević, Ciganović, and Tankosić—had organized the plot.

The highest level involvement of Serbia stands as evidence of the fact that Narodna Odbrana agents reported their activities to their president, Boža Janković, who in turn reported to the Serbian prime minister Pašić. The Narodna Odbrana as a terrorist organization could not be maintained with its headquarters in Serbia if the Serbian government would not accept its goals. None would risk a diplomatic row, economic and political sanctions, and isolation as those imposed against Yugoslavia, in fact Serbia, between 1991–1999, just for the sake of maintaining an organization if it had no serious intentions to use it.

This presumption is further strengthened by Pašić's refusal of extradition of the three men accused of masterminding the plot. He told the Austro-Hungarian government that he was unable to hand over the three men as it "would be a violation of Serbia's Constitution and criminal law." At the Sarajevo trial the three assassins from Belgrade were instructed to try to take all blame on themselves. Čabrinović and Princip claimed that the idea of the heir came from a newspaper clipping announcing Franz Ferdinand's planned visit to Sarajevo. The defendants refused or were unable to provide details under cross-examination, which meant that they were instructed to do so.

The Narodna report evidences of another conspirator from the military, Serbian Major Kosta Todorović, listed by the Austro-Hungarian Redbook as boundary commissioner and director of Serbian military intelligence services for the frontier line from Rada to Ljuboija in 1913. The Austrians later confiscated the Narodna report, as well as Pašić's handwritten notes, and additional incriminating documents. Pašić's handwritten notes from the briefing, estimated to have taken place on June 5, 1914, included the nickname of one of the assassins ("Trifko" Grabež) and also the name of Major Tankosić.

Further evidence is that during the 1917 Saloniki trial, the Serbian prosecutor cut off the speakers if they were speaking about the Sarajevo assassination. Serbia published no clarification of their confessions with regards to the Sarajevo attack, because it would have been incriminating; therefore, as usual it executed the hired executives. At trial it was forbidden to talk about Sarajevo, in spite of or exactly because of that, and the trial and execution of the masterminds was exactly staged and done because of the prospects of a détente, which seemed to be the best solution to alleviate Serbia's defeat in World War I. This was obvious in 1917, when the whole county was under Austro-Hungarian occupation and its government was being exiled; therefore, the trial was held in French-held Saloniki, in Greece. Pašić started to fear that the perpetrators may tell of his involvement and full knowledge of the plot to a court and execution of the perpetrators in 1917, after a mock trial.

Due to the suppression of Apis's confession and of the Saloniki trial transcripts by Serbia, Rade Malobabić was initially not closely linked to the Sarajevo attack. He arrived in Sarajevo on the eve of the attack and gave the final go ahead for the operation to Danilo Ilić. In 1914, Malobabić was the Serbian military intelligence's chief undercover operative against Austria-Hungary, as revealed by Serbian documents captured by Austria-Hungary during the war. They describe the provision of arms, munitions, and agents from Serbia into Austria-Hungary under his direction.

Another officer, Ilić, recruited the Serbian teenagers Vaso Čubrilović and Cvjetko Popović shortly after Orthodox Easter on April 19, 1914, for the assassination, as evidenced by their testimony at Sarajevo. Three teenagers—Princip, Grabež, and Čabrinović—were Austro-Hungarian citizens; in fact Bosnian Serbs living in Belgrade, who testified at Sarajevo after Easter said that they were eager to carry out an assassination. They also testified that Tankosić, directly and through Ciganović, provided the weapons and ammunition. The four pistols supplied to the assassins were found, having serial numbers 19074, 19075, 19120, and 19126. Princip used pistol 19074. Tankosić also provided training, a special map with the location of gendarmeries marked, contacts on a special channel used to infiltrate agents and arms into Austria-Hungary, and a small card authorizing the use of that channel. He also provided money and suicide pills in case the perpetrators would be caught. Tankosić confirmed to historian Luciano Magrini that he provided the weapons and was responsible for training the assassins and he initiated the idea of the suicide pills.

The planning of a direct action against Austria-Hungary came in late 1913, when Ilić came to a listening post at Užice to speak to the officer in charge, Serbian colonel C. A. Popović, to recommend it. Popović passed Ilić on to Belgrade to discuss this matter with Chief of Serbian Military Intelligence Colonel Apis. During this January 1914 meeting, various possible Austro-Hungarian targets for assassination were discussed, including Franz Ferdinand. Soon after their

meeting, Apis's right-hand man Tankosić, called a Serbian irredentist planning meeting in Toulouse, France, where they enrolled Mehmed Mehmedbašić, the son of an impoverished Muslim noble from Herzegovina.

The conspirators traveled along the Sava River to Šabac where they contacted Popović, who in turn, passed them further to Serbian Captain Prvanović, and filled out a form with the names of three customs officials assassins to receive discounted train tickets for the ride to Loznica, a small border town. After reaching Loznica on May 29, Prvanović summoned three of his sergeants to discuss the best way to cross the border undetected. Čabrinović handed over the weapons to Princip and Grabež. Princip told Čabrinović to go alone to Zvornik, to make an official crossing there using Grabež's identification card and then go on to Tuzla and link back up.

On May 30 Prvanović's sergeants led Princip and Grabež with the weapons by foot to Isaković's Island, a small island in the middle of the Drina River on the border. A Sergeant Grbić passed the terrorists into Austro-Hungarian territory from agent to agent, from safe house to safe house in the Narodna Odbrana network. After weeks of hiding, Ilić handed over the weapons only the day before the assassination, on June 27, 1914. Until that day he had kept secret the identities of the hit men from Belgrade, as well as those he had recruited locally. On the morning of June 28, Ilić walked on the street from assassin to assassin encouraging all seven of them to bravery.

Princip killed Franz Ferdinand on the Appel Quay, near the Cumurija Bridge. Franz Ferdinand hadn't learned from the case of Julius Caesar, who didn't wanted to go to the senate on March 15, 44 BCE, but to overcome rumors about his bad health, he ignored a report handed to him warning him about the impending assassination plot. Anti-Serb rioting broke out in Sarajevo in the hours following the assassination until order was restored by the military. This showed Bosnian sympathy for Austria-Hungary rather than for Serbia, and history repeated itself in this sense as well during the 1992–1995 Bosnian wars, when Bosnians defended heroically their capital, Sarajevo, from the Serbian siege, which lasted three years.

The adult defendants, facing the death penalty during the October 1914 Sarajevo trial, portrayed themselves as unwilling participants in the conspiracy. Veljko Cubrilović, a Narodna Odbrana agent, stated that he was acting from fear of a revolutionary organization capable of committing great atrocities known to exist in Serbia and also because Princip warned him "If you betray us, you and your family will be destroyed." Cubrilović acted because "he was more afraid of terror than the punishment of the law." Princip said that he was a Serbian nationalist who wanted to free his country "by means of terror."

The court was acting in full accordance with the Austro-Hungarian law, sentencing the perpetrator Gavrilo Princip to 20 years in prison as a minor, the others, Čabrinović and Grabež, got lighter sentences on the same grounds. Only

three adults—Ilić, Čubrilović, and Jovanović—were sentenced to death by hanging; the rest of the defendants got lighter charges up to 10 years in prison or were acquitted as a sign of Austro-Hungarian forgiveness.

The Weaknesses of the Alternative Theory of the Black Hand

An alternative theory to the Sarajevo attack states that it wasn't as much the Serbian military intelligence but the Black Hand that organized the plot. To show the weaknesses of this theory, it should be noted that the Black Hand was formed by people in the Serbian military to counterbalance the impact of the Internal Macedonian Revolutionary Organization, which was backed by Bulgaria. However, the Black Hand was ineffective in the end because its former president died and wasn't replaced, so it had no leadership, no active secretary. The casualties of the Balkan War and its aftermath caused disorganization within the Black Hand, and by early 1914 it was operating almost completely at the behest of Apis. The conspiracy under the name Black Hand also penetrated the Narodna Odbrana in Bosnia. The overlap in membership between the Serbian military and the Black Hand was intentional to make most evidence ambiguous in case it would be discovered and to avoid responsibility.

Apis confessed that he ordered the operation in his capacity as chief of the general staff. As such, it was not at the impetus of the Black Hand itself, as the organization's constitution, which required assassinations to only be carried out on a vote by its Supreme Council, was not in effect. Although some Black Hand leaders were involved, this does not disprove official involvement, since precise administration or recording was and is still not a Balkanic feature. The chain of command was simple: the intelligence department → Black Hand member officers → assassins. This is a widespread path of action around the world, where secret police formations have killed political dissidents abroad this way. The refusal of extradition of Milan Ciganović resembles that of the present day protection of Bosnian war criminals Ratko Mladic and Radovan Karadžić, whose extradition the international community has requested since 1996. Recently the psychiatrist Karadžić was allowed "to be found" in Belgrade in July 2008 in order to ease the pressure mounting on Serbia for a decade, but Colonel General Mladic is still on the run.

References and Further Reading

Albertini, Luigi. *Origins of the War of 1914*, Vol. II. London: Oxford University Press, 1953.

Belfield, Richard. *The Assassination Business: A History of State-Sponsored Murder*, New York: Carroll and Graf, 2005.

Dedijer, Vladimir. *The Road to Sarajevo*. New York: Simon and Schuster, 1966.

Evans, Robert John Weston. "Nationality in East-Central Europe: Perception and Definition before 1848." In *Austria, Hungary, and the Habsburgs: Essays on Central Europe, c. 1683–1867*. New York: Oxford University Press 2006.

Fay, Sidney Bradshaw. *Origins of the Great War*. New York: The MacMillan Company 1928.

Joll, James, and Gordon Martel. *The Origins of the First World War*. London: Pearson/Longman, 2007.

Juhasz, J., L. Markusz, P. Talas, and L. Valki. *Kinek a békéje? Háború és béke a volt Jugoszláviában*. Budapest: Zrínyi Kiadó, 2003.

MacKenzie, David. *Apis: The Congenial Conspirator. The Life of Colonel Dragutin T. Dimitrijevic*. New York: Columbia University Press, 1989.

MacKenzie, David. *Black Hand on Trial: Salonika 1917*. Boulder, CO: Eastern European Monographs, 1995.

Magrini, Luciano. *Il Dramma Di Seraievo. Origini e responsabilita della guerra europa*. Milan: NP, 1929.

Owings, W. A. Dolph. *The Sarajevo Trial*. Chapel Hill, NC: Documentary Publications, 1984.

Ponting, Clive. *Thirteen Days*. London: Chatto and Windus, 2002.

Schelking, Evgeniĭ Nīkolaevīch, and L. Makovskī. *Recollections of a Russian Diplomat, The Suicide of Monarchies*. New York: Macmillan, 1918.

Stoessinger, John. *Why Nations Go to War*. Belmont, CA: Wadsworth, 2007.

Strachan, Hew. *The First World War Volume 1: To Arms*. Oxford: Oxford University Press, 2001.

Taylor, A. J. P. *The First World War: An Illustrated History*. London: Penguin, 1963.

Treusch, Wolf Sören. *Erzherzog Franz Ferdinand und seine Gemahlin werden in Sarajevo ermordet*. Berlin: DLF, 2004.

Trydar-Burzinski, Louis. *Le Crépuscule d'une autocratie*. Florence: NP, 1926.

Živanović, Milan. *Solunski proces 1917*. Beograd: Savremena administracija, 1955.

CON

On the balmy Sunday morning of June 28, 1914, the Austrian Archduke Franz Ferdinand, the nephew of Franz Joseph I, Emperor of Austria and King of Hungary, who was visiting Sarajevo to review Bosnian military maneuvers, and his wife Duchess Sophie, were gunned down by a young Bosnian Serb named Gavrilo Princip, in league with several young Serb nationalists. Coincidentally it was also the Archduke and Duchess's wedding anniversary and, perhaps more ominously, Vidovdan, or St. Vitius Day, commemorating the Battle of Kosovo

in 1389 when the Serb lords were humiliated by the forces of the Ottoman Turks. Although the Second Battle of Kosovo in 1448 would play a more decisive part in the ascendancy of Ottoman rule in the Balkans, the 1389 defeat was a bitter one that future generations of Serbs would consider the defining moment of their collective identity.

While Princip and the other erstwhile conspirators could not have imagined the magnitude of their actions at the time, the ramifications of the murder in Sarajevo were cataclysmic; a spark that contributed to igniting the First World War and the last gasp for many long extant European empires. The identity of the assassin who cut short the lives of the Archduke and Duchess was never in doubt. However, the role of the Serbian government, if any, in the assassination continues to be debated amongst historians to this very day.

It is our position, as will be explicated below, that Princip acted alone and was neither directed, nor supported, by the Serbian regime. The young Bosnian-Serb was motivated solely by the nationalist passions of the day, which were further exacerbated by his own emotional instability. While he may have procured the assistance of some figures in the Serbian army, he acted in his individual capacity when he fired the shots that took the lives of the heir presumptive to the Hapsburg throne along with his wife, the mother of their young children. Further, it is unlikely that Princip was even acting in tandem with the nationalist groups with which he was associated at the time of the assassination.

Princip's Background

Gavrilo Princip was born in 1894 to a Bosnian Christian peasant family in Gornji Oblaj, a small and impoverished village in the outlying provinces. Princip left his home at the age of thirteen years and received his early education in Sarajevo, later completing his studies in Serbia. Already as a young man Princip had become acutely aware of the vulnerability of Bosnia to foreign rule, whether by the Turkish overlords or the Austrian dual monarchy. The nationalist movements that had previously emerged in Italy and Germany culminating in unification influenced Princip as well. He soon became a member of the Young Bosnian movement, a loose confederation of youthful nationalists, about which more will be said. If one's library and literary tastes provide an insight into one's political inclinations, Princip's personal collection of anarchist literature, including the works of Michall Bukanin and Peter Kropotkin, disclose a preoccupation with revolutionary thought and sympathy for armed terrorist crusades.

Historical Background

In order to place Gavrilo Princip and his political activities in context, it will be necessary to first establish the historical background that gave rise to the radical

nationalist and anti-imperialist rhetoric that encouraged and inspired young men like Princip to embrace a culture of politically motivated violence. In 1907 a series of international crises undermined the relative stability that Europe was then enjoying. Germany, who looked upon Austria as its sole reliable ally at the time, was compelled to allow the Austrians free reign in the Balkans. In 1908 the Austrian foreign minister Alois von Aerenthal met with his Russian counterpart, Alexander Izvolsky, and the two statesmen agreed that Russia would acquiesce to Austria's annexation of Bosnia and Herzegovina. In return, Austria would support Russia's efforts to gain access to the Turkish Straits enabling Russia to send its battleships through the waterway.

Without waiting for the Russians to benefit from their stake in the agreement, in October 1908, Vienna unilaterally announced its annexation of Bosnia and Herzegovina. Needless to say, the Russians were infuriated, and the Serbs, who themselves had hoped to take possession of Bosnia and Herzegovina, were similarly outraged. As the crisis spiraled, Germany pledged its unconditional support to its confederate Austria, forcing the Russians to back down. The Russians, still reeling from defeat in the 1905 Russo-Japanese War, were hardly in a position to aid the Serbs. Both their army and navy had proven incompetent against

Gavrilo Princip, suspected assassin of Archduke Franz Ferdinand, is hustled into custody in Sarajevo in 1914. He was a member of the Black Hand Secret Society which sought to create a greater Serbia through violence. (The Illustrated London News Picture Library)

the forces of the Empire of Japan. The humiliating loss had ignited a revolution at home.

In 1911 Italy went to war against the then ailing Ottoman Empire seeking to annex Tripoli on the Mediterranean coast of north Africa. The Turks were soundly defeated and a peace treaty signed in 1912 awarded Tripoli to the Italians. In the aftermath of Italy's swift victory, several small Balkan nations were emboldened to press their demands against the Ottoman Empire. Under Russian patronage, Bulgaria, Serbia, Montenegro, and Greece had formed the Balkan League. In 1912, the Balkan League launched a successful military offensive against Turkey. Under the terms of the Treaty of London (May 1913), the defeated Ottoman Empire lost all territory in Europe except for the area immediately adjacent to the Turkish Straits.

As this First Balkan War drew to a close, both Austria and Russia intervened. The Russians supported Serbia's demand for access to the Adriatic Sea, while the Austrians urged the formation of a new Balkan state, Albania, to contain Serbia's expansion. An international conference held in London in 1913 agreed to Austria's demand. The creation of Albania represented not only a setback for Serbia, but for Russia, as well. Once again the Russian role in determining the future of the Balkans was marginalized.

Having been denied access to the Adriatic, Serbia demanded part of Bulgaria's share of Macedonia as compensation. Bulgaria steadfastly maintained that its role in the First Balkan War entitled it to even more of Macedonia. These internecine disputes amongst the Balkan states eventually led to the outbreak of a Second Balkan War in 1913. Serbia, Montenegro, Greece, Rumania, and Turkey joined together to defeat Bulgaria. At the conclusion of the conflict, the Treaty of Bucharest (August 1913) required Bulgaria to cede land to Rumania, while Serbia and Greece gained most of the Macedonian territory.

While the crises in the Balkans had been serious, they did not lead to more wide spread hostilities. However, the post war political situation remained precarious. Russia's ambitions in the Balkans and the Turkish Straits had been frustrated. The Czarist government felt compelled to support Serbia more effectively in the future, while the Kaiser proclaimed Germany's unerring backing of Austria.

An eager and inspired Gavrilo Princip had volunteered for Serbian military service in the Second Balkan War, but his offer of service was declined by a recruiting officer because the young man was "too small and weak." The rejection was like a dagger in Princip's heart which, together with his animosity toward the Hapsburg Dual Monarchy, produced a profound resentment and hatred toward the Serbian military establishment.

Emerging of Nationalist Groups

Many student groups emerged in the aftermath of the Balkan Wars championing the nationalist ambitions of young Serbs such as Princip. The aims of these

groups were definitely not monolithic, but they did unanimously envision an independent Yugoslavia, a South Slav autonomous state, and an immediate realization of that vision.

Foremost among these youth movements was *Narodna Odbrana*, or the National Defense Society, which had been organized in 1908 in response to the annexation of Bosnia-Herzegovina by the Austrians. The ideology of *Narodna Odbrana* was predicated on vehement opposition to Austrian subjugation and the promotion of a strong sense of Serbian national identity. Training its members in paramilitary techniques was likewise a significant aspect of the Society's program. Recruits from Serbia as well as Bosnia-Herzegovina and Croatia were enlisted to its ranks. The stated aims of *Narodna Odbrana* were ostensibly modified in 1909 when the Society alleged its commitment to peaceful coexistence with Vienna. Consequently, the movement appeared to espouse a more culturally orientated agenda. The reality was, however, that *Narodna Odbrana* remained a belligerent and radical force. Indeed, in 1911, a National Defense Society publication originating in Belgrade asserted, in pertinent part, that "while the *Narodna Odbrana*, evolving with time, has adjusted itself to changes in the situation, it continues to maintain the links formed at the moment of the annexation and remains the same as it then was. Today demands quiet work, fanatical, indefatigable, and incessant, on the tasks and duties required as a preliminary preparation for the fight with rifles and heavy guns."

Under the aforementioned circumstances, the Austrian government was able to compile convincing evidence of *Narodna Odbrana* terrorist activities in Bosnia. The Austrians concluded that the Society was responsible for the murder in Sarajevo and made this formal charge in the court verdict. Since then, however, new evidence has come to light that indicates that the *Narodna Odbrana* was not responsible for the crime, or at the very least, not solely or directly culpable.

A secret Serbian organization which had infiltrated the *Narodna Odbrana* and used National Defense Society to further its own political agenda proved a willing and effective facilitator of the conspiracy. This association was called *Ujedinjene ili Smrt* or "Union or Death," more popularly known as the "Black Hand." The Austrian court charged with the prosecution of the assassins implicated certain "military circles" in Belgrade in its indictment. However, the court was not then aware of the existence of the Black Hand which these so-called "military circles" had organized and which was clearly implicated in the murder.

Most of the initial organizers of the Black Hand had been soldiers in the 1903 revolution that installed Peter Karageorgevich on the Serbian throne. Before the Balkan Wars, the Black Hand had promoted the Serbian cause by organizing armed bands and circulating propaganda materials in Macedonia. In Bosnia it had operated through the *Narodna Odbrana* and a nationalist group known as *Mlada Bosna*, "Young Bosnia." Revolutionary cells modeled on the Black Hand were established in Bosnia and controlled by the Central Committee located in Belgrade.

The new generation of South Slavs represented the backbone of "Young Bosnia": an amorphous nationalist movement rather than a clearly defined political organization. Its adherents did not constitute a single homogenous body. Rather, they were planted in various and diverse youth groups including agricultural cooperatives, choral societies, and numerous student organizations. However, they did share a common goal—liberation from Hapsburg rule and the introduction of sweeping social reforms. What they could not agree on was a course of action that would concretize these objectives.

While the aims of these nationalist groups were diverse, Princip and his colleagues eagerly embraced the idea of Yugoslavia as a South Slav independent state, and rejected gradualism and reformism as a viable means to achieving the desired end. Violence, they reckoned, would surely provoke further Austro-Hungarian repression and thereby increase South Slav hatred of the Hapsburg regime. Additionally, the employment of terrorism, tyrannicide, direct action—a decisive armed role by the individual actor in history—appealed to the radical Young Bosnians as well.

An Austrian diplomat described the atmosphere on the eve of the First World War as follows: "All who knew the country had the impression that an explosion was imminent. Especially in the schools, Pan-Serbian propaganda had created such chaotic conditions that a regular continuation of instruction hardly seemed any longer possible. Not only in Serbia itself but also in Austria-Hungarian territories inhabited by Southern Slavs, the conviction took root that the collapse of Austria-Hungary was near."

To be sure, the Black Hand had infiltrated the *Narodna Odbrana* and other nationalist groups, but its existence was not widely known outside government circles. It included in its membership an important contingent of the Serbian military and its operations were known to some foreign governments, but there was no direct link between the Serbian government and the Black Hand, per se. Colonel Dragutin Dimitrijević, its leader, was the chief of intelligence in the Serbian army, a figure of striking physical build, a monomaniacal Pan-Serb nationalist, of overpowering brute strength, he was codenamed "Apis" after a divine bull in ancient Egyptian mythology. Dimitrijević had been the mastermind behind the strategy that led to Serbia's successes in the Balkan wars.

According to a published account that first appeared in 1924, Apis had told a friend that Major Vojislav Tankosić, Apis's right hand man, had complained to him that several Bosnian youths were eager to perform some "great deed" on behalf of their country. These young patriots were aware that Franz Ferdinand was scheduled to review Bosnian troop maneuvers in June, 1914, and they wanted permission to go to Sarajevo. Tankosic later insisted that he opposed the request. But Apis responded favorably and encouraged Tankosic to give them a chance. After further reflection Apis became convinced that it was essential to kill the archduke. But he was unsure of the youths' ability to successfully effect

the assassination and tried, unsuccessfully, to restrain Princip to yield in favor of a more seasoned candidate for the job. Nevertheless, Gavrilo Princip was determined to go forward and he ignored Apis' request.

A month before the assassination of Franz Ferdinand, Princip had set out from Belgrade for a meeting with his fellow conspirators in Sarajevo. In order for Princip to cross the border between Serbia and Bosnia, he required the assistance of sympathetic agents, which he was given along the route, a route that had originally been laid out by *Narodna Odbrana* and was later used by the Black Hand. Once in Sarajevo, Princip met with his co-conspirators where they prepared the attack for a month hence.

Two members of the *Ujedinjenje ili Smrt*, acting under the cover of *Narodna Odbrana* and thereby eluding the detection of Austrian intelligence, played key roles: the aforementioned Major Vojislav Tankosić who provided the four revolvers and six bombs with which the group was equipped and Captain Rade Popovic who commanded the Serb guards on the Bosnian frontier and had seen Princip and his associates safely into Bosnia from Serbia some four weeks earlier. But the key figure behind both officers, and the driving force in *Ujedinjene ili Smrt* was Colonel Dragutin Dimitrijovic—Apis. His primary objective, and that of his organization, was not the federal Yugoslavia favored by the Young Bosnians but a Greater Serbia, with the expectation that the Serbs would dominate the Croats and Slovenes in the new state: 'the national ideal: the union of all Serbs.' At the time, Apis was in the grip of a fierce conflict with 68-year-old Prime Minister Nikola Pasic, a veteran politician who, like Apis, was a Serb nationalist, but unlike Apis, Pasic was cautious in his pursuit of Serb independence.

Apis and his claque in the Black Hand saw Pasic as an impediment to their goals. The prime minister's measured approach to Serbian autonomy was perceived by Apis and his like-minded military comrades as nothing short of treason. In 1914 the generals had forced Pasic out of office. While he was reinstated in short order, it required the intervention of the Russian and French governments on his behalf. This constituted a defeat for the bull-like Apis, necessitating some act of vengeance. He may well have considered an assassination of the Austrian Archduke a means of precipitating the collapse of the Pasic government.

Relationship between Serbian Government and the Black Hand

The relationship between the Serbian government and the Black Hand was a contentious one. Belgrade wished to avoid an armed confrontation with Austria which conflict, they surmised, would necessarily be instigated by the provocation of nationalist groups such as the Black Hand. In addition, Premier Nikola Pasic and his supporters espoused a pan-Serbianism rather than the Yugoslav ideal that inspired the revolutionaries. Following the Balkan Wars and the newly acquired Macedonian territories under Serbian rule, the government of Pasic and

the military authorities broke ranks over the administration of these lands. The Black Hand sided with the military, which explains in part why the latter was available to assist the Black Hand in its homicidal design.

The Black Hand has been characterized by some historians as an autonomous state in its own right. To be sure, Premier Pasic feared their power while he was careful to keep his apprehensions close to his chest lest the revolutionaries target him: he was keenly aware of the Black Hand's intentions concerning the archduke. But it is clear from all available archival documents that Pasic did not facilitate or authorize the attack. And there is even supporting evidence revealing Pasic's unsuccessful attempts to thwart the assassination. Prior to the murder, Pasic had ordered the frontier guards to halt the would-be assassins from crossing the border, and he had instructed the Serbian ambassador in Vienna, Jovan Jovanovic, to alert the Austrian government to the inchoate plot, of which Pasic had become aware in late May, 1914.

Ambassador Jovanovic was himself an ardent Serb nationalist and he was not inclined to carry through the mission assigned to him by Pasic. Instead, the ambassador met with the Austrian finance minister, who was not in a position to act on the proffered information. Jovanovic was also intentionally vague, confiding in the Austrian minister that "some young Serb may put a live rather than blank cartridge in his gun and fire it." No further details were provided, even though Belgrade knew the names of the conspirators. In all fairness to Pasic, he could not have known then how his message was diluted in the transmission. On the other hand, in the summer of 1914 Serbia was preparing for an election in which the political future of Pasic would be decided. It would not have been prudent for Pasic's party to alienate a large voting block who would consider him a traitor to the nationalist cause.

Unwilling to risk war with Austria and fearful of the powerful Apis and his organization, Pasic debated the options open to him: stop the conspirators at the border or warn Austria of the impending assassination. His decision was to walk a fine line and he consequently directed Jovanovic to act with caution yet never providing Vienna with a detailed account of the planned murder.

More significant is the official Serbian army publication, *Piemont*, which ran several articles highlighting the upcoming visit of Franz Ferdinand and linked it to the memory of "May 29, 1903." (The latter a reference to the coup d'etat led by Apis against King Alexander Obrinovic of Serbia culminating in the assassination of the monarch on June 11, 1903.) In such a highly charged atmosphere it is not surprising that young, emotionally and politically immature figures such as Gavrilo Princip would be caught up in the nationalist frenzy.

Princip did not approach *Narodna Odbrana* for money or armaments. Instead he and his co-conspirator Nedeljko Chabrinovich contacted the Black Hand leadership who expressed its approval of the assassination plot. The intervention of the Black Hand enabled the two Bosnian-Serbs to cross the border

and arrive in Sarajevo where they linked up with other radical youth awaiting the Archduke's fateful visit.

The key figure in the planning and execution of the attack was the afore-mentioned Apis—Dragutin Dimitrijević. His nephew, Dr. Milan Zivanovich, later explained the Colonel's motivation for supporting Princip and his co-conspirators. The archduke and heir presumptive to the Hapsburg throne favored a policy known as "trialism" in dealing with the restless Southern Slavs. Trialism envisioned the creation of a third state within the Hapsburg Empire consisting of territories inhabited by the Croats, Slovenes and Serbs. The Archduke was con-vinced that such an arrangement would satisfy the national aspirations of the South Slavs and put an end to the threat of insurrection by the Serbs. In fact, the Hungarians were vociferously opposed to any change in the composition of the Dual Monarchy, and the Southern Slav nationalists rejected the idea, perceiving in trialism an end to their vision of an independent Yugoslavia.

Some historians argue that the Serbian nationalists, rather than the govern-ment in Belgrade, had received false information indicating that the Austro-Hungarian army was set to attack Serbia. They point to the military maneuvers in Sarajevo as proof the Austrians were preparing for an imminent assault. After the Balkan wars, everyone knew that Serbia was militarily exhausted and would require several years to regroup. Franz Ferdinand, they surmised, planned to take advantage of this weakened condition.

Still other historians have even speculated that the Russian government was complicit in the assassination. This theory must be rejected in light of the fact that Franz Ferdinand was the leading Russian proponent in the Habsburg government.

Shortly before the June 28th murder, Serbian Prime Minister Nicola Pasic received an anonymous letter suggesting that the Austrian government might attempt to murder the archduke, who was never a beloved figure in the Habs-burg inner circle, particularly after his marriage to Sophia, which had been vehemently opposed by the emperor. The benefit, the anonymous writer sug-gested, would be that Austria could blame the Pasic government for the murder and commence military actions against Serbia.

Austrian Government's Response to Assassination

In Vienna the Austrian government responded to the murder by sending a legal expert to Sarajevo to collect evidence aimed at securing a successful prosecution. In July 1914 five of the assassins were placed on trial, the sixth defendant having managed to escape to Serbia. The defendants consisted of Princip, who commit-ted the murder, Chabrinovich, who had tossed the bomb that landed on the hood of the automobile in which Franz Ferdinand was driven, and three of their accomplices. The most vociferous and unremorseful member of the group was Princip, who boldly articulated his reasons for committing the murder: "I have

no regret," he said, "because I am convinced that I have destroyed a scourge and done a good deed . . . I have seen our people going steadily downhill. I am a peasant's son and know what is happening in the villages. . . . All this had its influence on me and also the fact of knowing that he [the Archduke] was a German, an enemy of the Slavs. . . . As future Sovereign he would have prevented our union and carried out certain reforms which would have been clearly against our interests."

After protracted hearings the court found the defendants guilty. The Kingdom of Serbia and the military intelligence service were deemed aiders and abettors in the crime. The verdict read in part:

> "The Court regards it as proved by the evidence that both the *Narodna Odbrana* and military circles in the Kingdom of Serbia in charge of the espionage service, collaborated in the outrage . . . There is no doubt that both the *Narodna Odbrana* and the military circles on the active list in the Kingdom of Serbia knew of the aims of the outrage and were prodigal of all possible assistance and all possible protection to the perpetrators for whom they actually procured the means of carrying out the assassination." (Albertini 1953: 68)

To be sure, the tribunal was correct in its assertion that military circles in the Kingdom of Serbia were, to a degree, conspired in the assassination. But the court erred in tying these military circles to the *Narodna Odbrana* and, by extension, finding the government of the Kingdom of Serbia complicit in the murder of Franz Ferdinand.

In conclusion, the extent to which the government of Serbia was behind the assassination of Archduke Franz Ferdinand has been argued by scholars for almost a century. There has not been one shred of evidence that the Kingdom of Serbia either instigated or orchestrated the attack in Sarajevo. At best, one may infer from the message Prime Minister Nicola Pasic attempted to transmit to Vienna that his regime was aware that some sort of conspiracy was afoot. But had his regime planned the demise of the Archduke, he most certainly would not have then sought to undermine the plot by informing the Austrians of its existence.

One might contend that Apis and the Black Hand were part of the Serbian government by virtue of the fact that Apis was the head of Serbian military intelligence. This argument cannot be sustained. The relation between the Black Hand, in general, and Apis, in particular, concerning the Pasic administration was, to say the least, fraught with tension. And more to the point, on June 14, 1914, Apis reported to a meeting of the Black Hand about the plan for Sarajevo a fortnight later. The committee members surprised Apis by rejecting the plot out of fear that such an act would inevitably lead to war with Austria, a position consistent with Pasic's. Subsequently, Apis tried through intermediaries to reach

---

The Fate of the Black Hand

Nearly four-and-a-half years after the assassination, World War I ended and the Austro-Hungarian Empire had collapsed. Sarajevo became a part of the new kingdom of Serbs, Croats, and Slovenes, which later became the Kingdom of Yugoslavia. Of the original group of members of the Black Hand who carried out the assassination, only two had survived: Mehmed Mehmedbasic and Vaso Cubrilovic, the former having escaped and the latter having been released at the end of the war.

Cubrilovic became a hero in Yugoslavia and went on to become a teacher and then a university professor in Belgrade. After World War II, he was appointed as minister of forests in the federal government. Another member of the Black Hand, although not involved in the actual assassination itself, was Ivo Andric, who was Yugoslavia'a ambassador to Germany when the Germans invaded in 1941. He went on to write the famous *Bosnian Chronicle* and was awarded the Nobel Prize for Literature in 1961.

The Yugoslav government in the 1920s and 1930s did much to commemorate Princip's actions, even naming the Latin Bridge, close to the site of the assassination, Princip Bridge. A set of footsteps on the ground marked the movement of Princip just before he shot Franz Ferdinand. Winston Churchill famously wrote in *The Unknown War* (1932), "Princip died in prison, and a monument erected in recent years by his fellow countrymen records his infamy and their own."

---

the relevant parties and have them abandon the scheme. But, despite his influence the directive was ignored.

Midway through the Great War, as Gavrilo Princip lay dying of tuberculosis in an Austrian prison, he was interviewed by psychiatrist to whom he declared that in going to Sarajevo "he only wanted to die for his ideals" and while he had been happy to accept the Black Hand's weapons he was not willing to obey the instructions not to use them. The smoking gun is clearly in Princip's hand and regardless of who may have procured the weapon, the young Bosnian Serb acted of his own volition and without any ties to the Serbian government.

References and Further Reading

Albertini, Luigi. *Origins of the War of 1914* (Vol II). New York: Oxford University Press, 1953.

Black, Jeremy. *Warfare in the Western World, 1882–1975*. Bloomington: Indiana University Press, 2002.

Di Scala, Spencer M. *Twentieth Century Europe: Politics, Society, and Culture*. New York: McGraw-Hill, 2003.

Donia, Robert J., and John V. A. Fine, Jr. *Bosnia and Hercegovina*. New York: Columbia University Press, 1995.

Goodwin, Jason. *Lords of the Horizons: A History of the Ottoman Empire*. New York: Picador, 2003.

Hamilton, Richard F., and Holger H. Herwig, eds. *The Origins of World War I*. New York: Cambridge University Press, 2003.

Keegan, John. *The First World War*. New York: Vintage, 2000.

Meyer, G. J. *A World Undone: The Story of the Great War, 1914 to 1918*. New York: Delacorte Press, 2006.

Morton, Frederic. *Thunder at Twilight: Vienna 1913/14*. Cambridge, MA: Da Capo, 2001.

Palmer, Alan. *Twilight of the Hapsburgs: The Life and Times of Emperor Francis Joseph*. New York: Atlantic Monthly Press, 1997.

Remak, Joachim. *Sarajevo: The Story of a Political Murder*. New York: Criterion Books, 1959.

Tuchman, Barbara W. *The Proud Tower: A Portrait of the World before the War 1890–1914*. New York: Macmillan, 1962.

4

The deaths of over one million Armenians in Turkey were due to a Turkish government policy of genocide.

PRO   James Frusetta
CON   Cenap Çakmak

PRO

The Armenian genocide is contested history. Such historiographical debates are not rare, but few are so profoundly divisive and politicized. The history of the *Aghed* ("catastrophe") and its portrayal as a deliberate attempt by the Ottoman state to exterminate the Armenian people has become a core element in the identity of the Republic of Armenia and diaspora. The Turkish government and many Turks oppose what they perceive to be a one-sided and inflammatory portrayal of events in the final years of the Ottoman Empire that ignores Turkish suffering.

This political debate has spread to international arenas. Twenty-one countries officially acknowledge the Armenian genocide, with recent debates over recognition within the legislatures of the United States, Sweden, Bulgaria, the United Kingdom, and Israel. U.S. House Speaker Nancy Pelosi in 2007 advocated the passage of a bill officially recognizing the genocide. Several European states (most notably France) have criminalized denial of the Armenian genocide. Given that the European Parliament and Council of Europe both formally recognize the genocide, there are ongoing debates over the official Turkish position in relation to the country's application for membership in the European Union.

Historians critical of the use of the term *genocide* have suggested, quite rightly, that political recognition does not confer scholarly accuracy. While acknowledging the moral, political, and legal debates, the focus here will be on the historical evidence (and debates over this evidence). Scholarly discussion on the question has become particularly fruitful in recent years, as the broader study of genocide and ethnic cleansing has expanded. The Armenian massacres have thus received fresh interest from historians as a subject of scholarly inquiry. Given certain unique elements of the Holocaust, there is considerable interest in the perceived utility of the Armenian case in comparative and theoretical studies, and the result has been a flood of specialized monographs and comparative works released in the past 15 years. The majority of recent English-language work has broadly bolstered the position taken by many (but not all) historians that the Armenian genocide was an act of genocide undertaken by the Ottoman government.

83

At the same time, as noted below, scholarship has raised new questions about how to best understand these events.

In resolving whether these deaths represent a state-sponsored genocide, there are four key questions to address: Is it accurate to describe these events as "genocide"? What was the larger context of the Armenian massacres? Was the government deliberately responsible for the deaths? What was the ultimate intent of the deportations and attacks against civilians and disarmed military units? Even among historians who agree that the Ottoman state carried out a policy of genocide, there remain significant disagreement and debate about details. In this light, scholarly critiques of the genocide argument could and should be engaged where possible in a healthy dialogue.

Wide-Reaching Massacres Occurred 1915–1918

Just as both sides of the scholarly debate over the Armenian genocide are frequently motivated by contemporary politics, the use of the term genocide has political overtones. For historians, however, the question is whether or not these events can best be described in such a fashion. For example, were the events of such sufficient scope as to destroy (or potentially destroy) the Armenian nation?

Many historians argue that the events of 1915–1918 fit the legal definition set forth in Article 2 of the United Nations Convention on the Prevention and Punishment of Genocide: "genocide means any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group." Within this definition, the term genocide is warranted—but, as some critics note, the convention is an extremely broad definition. Although Article 2 sets a standard in international law, the Armenian genocide does not fit some scholarly definitions. For example, the Holocaust historian Steven Katz defines genocide as applying "*only* when there is an actualized intent . . . to physically destroy an *entire* group" (1994: 128). Similarly, the political scientist Guenter Lewy (2005) argues that there is insufficient evidence to demonstrate intent on the part of the Ottoman government to commit the complete destruction of the Armenian people. Both arguments reflect the continued problem of creating a universal definition of genocide for scholars and its utility in relation to similar terms such as democide, cleansing, or ethnocide. Despite its flaws, the definition as provided by Article 2 has the merits of international recognition and standing (albeit in legal and political circles) and thus serves as the threshold here.

What were the series of events that have come to be known as the Armenian genocide? Even before and in the first months of World War I, the Ottoman government led by the Young Turk movement (İttihad ve Terakki Cemiyeti, or Committee for Unity and Progress) was suspicious of Armenian loyalty and, some historians argue, had already taken steps that would lead to the massacres. For example, the paramilitary forces of the Special Organization (Teşkilatı Mahsusa)

were created before the Balkan Wars to deal with separatism and counterinsurgency operations. In August 1914 the Special Organization was expanded and, in subsequent months, was used to seize strategic border areas, securing them in some cases through attacks on civilians (including Armenians). Within the government, orders were given to monitor prominent Armenians and warnings were given that the group was subversive and disloyal. Such orders reflected that a separatist movement did exist, although the extent and support for such among Armenians remains in dispute. The presence of Armenian volunteers fighting alongside and within the Russian army (often Russian citizens themselves) was further considered evidence of disloyalty and potential danger.

Such concerns increased in December 1914 when the Ottoman invasion of the Russian Caucasus ended in disaster and with the near destruction of the Ottoman Third Army. Russian forces counter-invaded, a process that included attacks on Muslim civilian populations by irregular forces. The result was increasing fear that Armenian rebels were engaging in sabotage and guerrilla operations, enflamed by (false) accusations that Armenian "treason" had been responsible for the Ottoman defeat. In the spring of 1915, Armenian recruits in the Ottoman military were removed from combat units, disarmed, and placed in labor battalions (many of which would be subsequently executed). At the same time, the Special Organization, army, and police moved to confiscate weapons owned by Armenian civilians. Numerous incidents of assault, torture, and murder of civilians were reported, usually but not exclusively against adult men.

Arrests, local deportations, and massacres occurred on a relatively small scale into April 1915, when conflict erupted in the city of Van. Declaring the Armenian residents of the city to be in rebellion for offering to meet only part of a military conscription quota, the Ottoman army besieged the town, and open fighting erupted. Arguably as a security measure and in response to the events in Van and to the threat of Allied landings in Gallipoli, the government arrested hundreds of Armenian leaders in Istanbul (many of them intellectual figures rather than politicians) over April 24 and 25. April 24 has accordingly become the official day of remembrance for the event of the Armenian genocide, since these leaders would be first deported and the overwhelming majority executed months later. Policies of "punishment" for the Armenian population now gave way to policies intended to secure Anatolia from the threat of secession.

The Tehcir Law (the Temporary Law of Deportation) of May 1915 accordingly stipulated the removal of Armenians from "war zones" to a series of camps in the Syrian Desert. The policy was quickly extended to Armenians living in provinces behind the fronts (ironically enough, resulting in the transportation of Armenians from central Anatolia to Syria, closer to Ottoman lines against the British in Mesopotamia). Although rail transport by boxcar and cattle car was provided on some occasions, hundreds of thousands of civilians (including women and children) were forced to march all or part of the distance to the

relocation camps. Inadequate shelter, food, and water were provided, decimating the marchers; despite sporadic offers of aid (including from U.S. and German diplomats), the Ottoman government refused to allow foreign third parties to assist with the relocations by providing aid. Columns were subject to attacks by irregulars and bandits; others were attacked by Special Organization units or victimized by the guards assigned to the columns. Vahakn Dadrian (2003) argues that the camps themselves were used as sites for deliberate "mass killing." At the very least, the camps were a form of passive execution as a result of privations. The bulk of the deportations occurred over 1915–1916, but the process continued through World War I.

There is considerable debate over the existence and accuracy of contemporary evidence for the massacres. Contemporary British, French, and Russian accounts were undoubtedly colored by the desire to castigate the Ottomans with wartime propaganda. "Neutral" accounts (including that of U.S. Ambassador Henry Morgenthau) are sometimes criticized for not being eyewitness accounts. Some eyewitness neutral accounts (for example, by Protestant missionaries) are similarly questioned as showing bias toward the Christian population. Similarly, bias is ascribed to the accounts of Armenian survivors, often recorded decades after the events or posing methodological problems as oral sources. Although postwar trials by the Ottoman government provide a wealth of information, they were arguably influenced by political expediency and the need to placate the Allies and limit the severity of the imposed peace treaty. While stipulating such scholarly concerns, significant contemporary documentation does exist as written by Ottoman authorities as well as their wartime Austro-Hungarian and German allies (the latter including thousands of military advisers and attaches in Anatolia, some of whom condoned or advised on the deportations and massacres).

Given the preponderance of information, there is broad agreement that the deportations and massacres did occur, although there remains considerable debate over specific actions, sources, and the total number of casualties. It is possible to confirm the severity and scale of the massacres as being such as to warrant the scope of the term genocide. Ascribing such, however, raises the question of context and intent.

The Armenian Genocide Took Place in a Larger Context

One criticism frequently raised against defining the massacres as a state-organized genocide was the nature and severity of conflict in the region. This includes not only the interstate conflict of World War I, but also insurgency and counter-insurgency operations, banditry, renewed fighting in 1918–1923, and widespread famine and disease that corresponded with these conflicts. In this light, critics argue, Armenian deaths occurred in the context of "civil war" (rather than genocide) in which Turkish deaths and suffering are ignored. Such criticisms are well

---

Henry Morgenthau and the Ottoman Empire

One of the critics of the Turks during the war was Henry Morgenthau (1856–1946), the U.S. ambassador to Constantinople. Born in Mannheim, Germany, from a prominent Jewish family involved in the cigar industry, he immigrated to the United States with his parents when he was 10. He studied law and went to work for a law firm.

Becoming interested in labor rights and a supporter of Woodrow Wilson, he took up his position in the Turkish capital in November 1913. One of his concerns, with the outbreak of World War I, was that if the Ottoman Empire entered the war, it could cut off foreign access to the Jewish settlers in Palestine. However, when war started and the Ottoman Empire entered the war on the German side, Morgenthau found himself increasingly critical of the lack of action over the killing of Armenians and other Christians. Because of this he resigned his position and later was involved in a secret mission to mediate a separate peace between the Allies and the Ottomans. After the war, he became heavily involved in work for Greek refugees. His son, also called Henry Morgenthau, became the U.S. secretary to the treasury under Franklin Roosevelt.

---

taken in stressing the context of the events. Focusing too closely on the deportations and massacres of 1915–1918 neglects earlier anti-Armenian pogroms in the 1890s and 1900s, as well as continued conflict after 1918. Even more significantly, it neglects the sustained cleansing of Muslims (including but not exclusively ethnic Turks) from southeastern Europe and the Caucasus as former Ottoman territory was lost to the Russian Empire and to the independent states of Serbia, Greece, and Bulgaria from 1800 to 1923. Although resolving here that Armenian genocide itself did occur as a distinct process, it should and can be placed in a larger context that recognizes a larger process of massacre and cleansing against both Christians and Muslims.

Recent historiography has taken odds with the long-lasting ideas that the Ottoman Empire went into decline after 1699 and the persistent stereotypes denigrating its society, politics, and accomplishments. This does not mean there were not significant problems, notably the state crisis in the early 19th century known as the *kurdzhaliistvo* (Bulgarian for the "time of the brigands"), when central control over much of the empire collapsed and provincial military forces turned to banditry and warlordism. Local Muslim rulers (many not ethnically Turkish) contested for power with the sultan and the central government of the Sublime Porte, and both Muslim and Christian peasants suffered during the breakdown of governance. The subsequent decades of the 19th century were marked by continued policies of modernization of the military and the recentralization of authority in an attempt to address problems of inefficiency, corruption, and lagging technical development.

Efforts at reform were spurred by continued Russian pressure in the Caucasus and the growing demand by many of the empire's Balkan subjects for independence. The Russian Empire was the strategic victor in its eight wars with the Ottoman Empire over the 18th and 19th centuries, annexing much of the latter's territory in the Caucasus. The adoption of the nation–state ideal—the vision that states should represent an ethnic group—saw successive Serbian, Greek, Romanian, Bulgarian, and Albanian movements seek to end Ottoman suzerainty over their claimed homelands and transform the local multiethnic and multireligious society into homogenous nations. Secessionist trends were aided by foreign intervention. True, the European great powers (particularly Great Britain) did consistently act to prevent the final dissolution of the Ottoman Empire. But they also frequently intervened for their own commercial and political benefit or to support local uprisings—claiming the legal right to intervene on behalf of the sultan's Christian subjects. The loss of territories thus followed a pattern: The trigger was usually an uprising within the empire by one of the Christian populations (Muslim uprisings also occurred, but, with the exception of Bosnia in 1875, rarely prompted foreign intervention). Repression of the uprising, frequently brutal and including attacks by irregulars and local Muslim militias on civilians, prompted foreign intervention up to and including military force to secure specific rights for a local Christian population. Acknowledgment of local self-government and autonomy within the empire led, over time, to secession and the declaration of independence during a future round of Ottoman state crisis. In several cases in southeastern Europe, this pattern was recognized such that local revolutionaries arguably sparked confrontations with the Ottoman government they knew they would be unable to win, but they believed the resulting bloodshed would ensure great power intervention on their behalf.

As Justin McCarthy (1995) argues in *Death and Exile*, these successive losses of territory caused millions of Muslims to flee to Anatolia in the *Muhajiria* (from the Arabic *muhajir*, or "refugee"). Much of this flight occurred when Muslim populations privileged within an Islamic state were now incorporated within either an Orthodox Christian Russian empire or Christian nation–states in the Balkans and feared either a loss in status or persecution. Such fear was not unjustified, as massacres (and countermeasures) by Christian irregular forces occurred throughout these conflicts. The refugee waves were resettled within the empire and frequently were employed against new uprisings; *muhajir* groups participating in the Armenian massacres most likely *did* see these events as the continuation of earlier, anti-Muslim cleansings.

Successive political movements in the 19th century attempted to redefine the empire and address the concerns of its constituent population, including the non-Islamic minorities. The early Tanzimat reformers of the 1840s and the Young Ottomans of the 1860s hoped to modernize the empire and transform it into a parliamentary and increasingly secular state. Sultan Abdülhamid II's

encouragement of Pan-Islamism in the 1890s turned to attempts to revitalize the Islamic Caliphate and define the empire along strict Islamic lines. The failure of each movement to transform the empire and resolve core problems helped propel the Young Turks to power through two coups d'état in 1908 and 1913. To reform what remained of the Ottoman multinational empire, the Young Turks advocated the concept of the nation–state and the transformation of the Ottoman Empire into a Turkish state. The loss of Libya, Albania, Macedonia, and western Thrace in the Italo-Ottoman War of 1911 and the Balkan Wars of 1912–1913 provided fresh incentive, as the empire was now reduced to a core of eastern Thrace, Anatolia, and its restive Arab territories in the Middle East.

The Young Turks, with justification, saw the presence of non-Muslim (and now non-Turkish) inhabitants in the remaining territories as, in the words of Robert Melson, "threatening and alien minorit[ies]" (1992). As discussed below, the presence of minorities was perceived as threatening the remaining core of the empire. Taner Akçam (2004, 2006) has argued persuasively that the need to remove minority groups, including the Armenians, was discussed and debated openly by elements of the Young Turk movement prior to World War I. Although it is not necessarily conclusive that the genocide was planned, it is highly suggestive of a widespread perception that the empire's minority population was a crisis in waiting.

Given these concerns, it is important to note that during the Armenian deportations, similar actions were taken with respect to other minority groups, including both Greeks and Assyrians. Although the respective historiographical debates have not reached the prominence or intensity of that of the Armenian genocide, some historians have proposed that both cases should be treated as genocide (particularly with reference to the Assyrians, who suffered proportionally almost as much as the Armenians did). Even some Muslim groups such as the Kurds, Bosnians, and Albanians who were living within the empire were internally resettled, sometimes forcibly.

Finally, there remains the question of defining the ending point of the genocide. While much of the literature focuses on the events of 1915–1918, Ottoman wartime policies merged into the postwar conflicts in eastern Anatolia. Western historiography generally stresses the formal end of World War I on November 11, 1918. This neglects the continuity of conflict in much of Eastern Europe, as a series of not unrelated conflicts raged: civil wars in the former Russian Empire, the Baltic states, Finland, and Hungary; civil unrest in the kingdom of Serbs, Croats, and Slovenes; the wars of Soviet Hungary and the Greco-Turkish War. The collapse of the Russian Empire in the Caucasus resulted in the emergence of (briefly) independent states. Border clashes broke out between the new Armenian, Georgian, and Azerbaijani states, complicating the pattern of ethnic violence in the region (particularly with Soviet intervention). In 1918 the Democratic Republic of Armenia (formerly Russian Armenian, now briefly

Armenian widows and children, about 1915. (Library of Congress)

independent) launched an invasion of the Ottoman provinces of Van, Erzurum, and Trabzon in revenge for the Ottoman massacres, the attack tolerated by the postwar French Occupation Forces. Nationalist Turkish forces invaded Armenia in turn in 1920, invading Kars and Ardahan (and leading independent Armenia to acquiesce to Soviet takeover, in part for protection from invading Turkish forces). Such fighting is indeed close to the critical portrayal of the massacres as occurring within a context of "civil war" in which casualties were inflicted on all sides. This does not mitigate the events of the genocide, in that this phase of violence occurred significantly later. But, and in parallel with policies of cleansing and massacres of both Turkish and Greek civilians during the Greco-Turkish War of 1921–1923, it represents an aspect of the history of ethnic violence in this period that is frequently forgotten.

Recognizing that significant actions against Muslims took place—including events that can also be identified as genocide—does not diminish or excuse the Ottoman government's involvement in the Armenian genocide of 1915. It does, however, argue for the need for scholarly attention to the systematic massacre, deportations, and flight of Muslims along the Ottoman frontiers. The fear of fragmentation, given the existing pattern of minority separatism in the empire, was, arguably, a legitimate fear.

Massacres Were the Policy of at Least Segments of the Ottoman Government

Establishing that massacres occurred in an environment of violence and conflict begs the question to what extent the massacres were organized by the Ottoman government. Archival data detail involvement at both levels. In Istanbul, the Ministry of the Interior (headed by Mehmet Tâlât) and the Ministry of War (led by Enver Paşa) directed provincial governors to undertake the deportations. Although some wartime records were subsequently lost or destroyed, there remains a considerable correspondence on the part of the central government that details its involvement. In the field, units of the Special Organization, police, gendarmerie, and (more rarely) the army were tasked with rounding up and escorting the deportees, as well as disposing of the property the deportees left behind.

Here, again, critical assessments of the genocide make the valid point that the Ottoman state cannot be treated as a monolith. As with any state, the empire's government was fragmented between various factions, organizations, and institutions, and not all were equally involved in the massacres. In the same way, although Vahakn Dadrian argues for the root culpability of the Young Turk movement in planning the genocide (likening it to the role played by the Nazi Party in the German organization and carrying out of the Holocaust), there remained factions within the movement. This can be seen by the mixed reactions by elements of the state apparatus to the massacres. Some army leaders openly questioned or refused to be involved with the deportations (although others actively supported them). In several cases, provincial governors (such as Tahsin Bey in Erzurum, Mazhar Bey in Ankara, and Reşit Paşa in Kastamonu) refused to undertake the orders, questioned if they should be carried out within the province, or requested that the orders be limited to adult men and exclude women, children, the elderly, and the sick.

Critics have advanced, in some cases, the reverse argument—that in fact the gendarmerie and military units involved in the genocide took actions that were not approved by the state. Here, too, this raises the significant point that "wild" massacres and cleansings likely occurred (as in most cases of genocide or ethnic cleansing). But the duration of the events and the fact that the central government did not step in to stop them (even if hampered by the need to wage war simultaneously on three fronts) suggest that the government supported "wild" actions—which were, in any case, undertaken by state officials. The overall picture is complex, but it still remains a picture in which elements of the state ordered, carried out, and tolerated the deportations and massacres.

That said, was genocide the intent of the state, or did the deaths emerge out of ongoing conflict throughout the region? Many historians argue that Ottoman policy in the 1890s and 1900s represented antecedents that led to the wartime massacres— what Dadrian refers to as a "proto-genocidal" policy. In this argument, World War I

created conditions under which an existing ("intentionalist") policy could be carried out. Donald Bloxham (2005) gives an equally compelling "functionalist" argument in *The Great Game of Genocide*. Acknowledging existing nationalist sentiment and discussion, Bloxham argues that the crisis of World War I provided conditions for the radicalization of policy and sentiment and for the widespread toleration (even support) of the massacres by Ottoman officials and citizens. There is valid debate regarding at what point government policy became one of outright genocide. Even if, however widespread the casualties, the deportations were not originally intended as genocide, the Ottoman government did not halt operations or provide additional security, shelter, food, or transportation. This suggests that even if much of the Ottoman state had not been involved in planning or supporting genocide, ultimately tens of thousands of civil servants, police, and military personnel became complicit once the deportations began killing hundreds of thousands of people.

The Goal Was the Elimination of the Armenian People from Anatolia

At first hand, the logic of the 1915 massacres is elusive. Not only did the massacres deny the Ottoman Empire a source of manpower in a time of war (many Armenians having served bravely in the army in the campaigns of 1914), but it also created internal conflicts with what had until the late 19th century been referred to as the "most loyal *millet*," the Armenians. Critics rightfully question what motivation the empire could have had to sustain a lengthy and costly campaign of massacres and deportations. If the goal was to exterminate the Armenian people, then why did the state undertake deportations (however ultimately lethal) instead of immediate murder? Moreover, why were the Armenian communities in Istanbul and Izmir generally left intact (albeit with elements of their leadership imprisoned and executed)?

Past service to the Ottoman state by the Armenian *millet* was now overshadowed by recent events, such as the instance of an Armenian delegation presenting a petition in 1877 to the British and Russians, wherein they demanded extensive changes (including autonomy) in the Armenian-inhabited provinces. In the context of the 19th century, the pursuit of autonomy was normally a step toward an independent state and the secession of territory. An independent Armenian state (and Greek claims on Greek-inhabited areas of the Anatolian coast and eastern Thrace) would effectively divide the core of the empire. And certainly, there were movements among the Armenians (although of disputable size, and almost certainly not extending throughout the entire population) that sought autonomy. In this light, the logic of successive governments' increasing portrayal of the Armenians as "traitors" and as scapegoats for the problems of the society can be understood (if not condoned). Major massacres in 1895–1896

and 1909, in which tens of thousands of Armenians were killed, were thus "collective punishment" for the group's perceived treason.

As noted above, before World War I, there were considerations by Young Turk leaders of the need to remove minorities from strategic border zones—not only the Armenians, but also Greeks and others. At the beginning of the war, tens of thousands of Greeks were forced to evacuate the coast and were either deported internally (in marches not unlike those of the later Armenian massacres) or encouraged to immigrate to Greece proper. This expanded to the perceived need to remove minorities that could threaten the secession of territory, most notably the Armenians.

This did not necessitate the complete annihilation of all ethnic Armenians—although the deportations came close to achieving this. It did require the elimination of the Armenian people *as a group* from eastern Anatolia. Individual Armenians could survive through conversion to Islam and assimilation; strikingly, Taner Akçam (2004, 2006) argues that Ottoman officials saw no need to deport or eliminate Armenians from areas where their population was smaller and less dense. This fits, in his arguments, a broad policy in which *all* non-Turkish minorities (including other Muslim groups) were to be relocated to ensure that minorities remained less than 5 percent of a given provincial population. This explains the relative preservation of Armenian communities in western Anatolia that were likely viewed as nonthreatening enough to warrant deportation (particularly given the larger number of foreign observers in those regions who would have witnessed the events firsthand).

In this sense, the Armenian genocide was undertaken to remove a perceived threat to continued Muslim, Turkish, or Ottoman rule over the eastern regions of Anatolia. Unlike the Holocaust, the Young Turks did not perceive the question in a racial sense, nor did they take it to the extreme of defining the Ottoman Empire's problems in a monocausal fashion. For example, while Armenians were frequently used as scapegoats for the state's problems, they were not (as the Nazis did with the Jews) portrayed as a universal and global conspiracy undermining civilization. As long as the nation was removed or eliminated, individual Armenians or families might be tolerated if they were assimilated into Muslim society.

This argument does not fully explain the brutality and bloodshed of the deportation and massacres, because nonviolent removal could have achieved most of the same goals. Scholars have accordingly raised relevant questions regarding the possible desperation felt by state officials, the depth of extreme nationalist anti-Armenian sentiment among planners, the extent of control over the deportations, and the larger questions of violence, anger, and ethnic conflict in Ottoman society in eastern Anatolia. Stipulating that the goal of the Armenian genocide was the removal of the Armenian people from Anatolia does not fully explain the extensive violence and brutality, but it does fit the definition of genocide.

Conclusion

Given these four issues, it is resolved here that although the events in question remain complex and there are valid academic debates over the details, the death of over 1 million Armenians was the result of Ottoman state policy. Reasonable historical evidence exists that the massacres took place, that they were the design (or outgrowth) of government policy, and that they were undertaken for the specific purposes of defending the state—albeit in a complex period in which the Armenian genocide was one part of a larger mosaic of ethnic cleansings, massacre, and violence.

References and Further Reading

Akçam, Taner. *A Shameful Act: The Armenian Genocide and the Question of Turkish Responsibility*. New York: Metropolitan Books, 2006.

Akçam, Taner. *From Empire to Republic: Turkish Nationalism and the Armenian Genocide*. New York: Zed Books, 2004.

Bloxham, Donald. *The Great Game of Genocide*. Oxford: Oxford University Press, 2005.

Dadrian, Vahakn. *The History of the Armenian Genocide*. New York: Berghan Books, 2003.

Hannibal, Travis. "Native Christians Massacred: The Ottoman Genocide of the Assyrians During World War I." *Genocide Studies and Prevention* 1, 3 (2006): 327–371.

Hovannisian, Richard. *The Armenian Genocide*. New Brunswick, NJ: Transaction Publishers, 2007.

Hovannisian, Richard. *Remembrance and Denial*. Detroit, MI: Wayne State University Press, 1998.

Katz, Steven T. *The Holocaust in Historical Perspective: The Holocaust and Mass Death before the Modern Age*. New York: Oxford University Press, 1994.

Levene, Mark. "Creating a Modern 'Zone of Genocide': The Impact of Nation- and State-Formation on Eastern Anatolia, 1878–1923." *Holocaust and Genocide Studies* 12, 3 (1998): 393–433.

Lewy, Guenter. *The Armenian Massacres in Ottoman Turkey: A Disputed Genocide*. Salt Lake City: University of Utah Press, 2005.

McCarthy, Justin. *Death and Exile: The Ethnic Cleansing of Balkan Muslims*. Princeton, NJ: Darwin Press, 1995.

Melson, Robert. *Revolution and Genocide*. Chicago: University of Chicago Press, 1992.

Naimark, Norman. *Fires of Hatred: Ethnic Cleansing in Twentieth Century Europe*. Cambridge, MA: Harvard University Press, 2001.

Reid, James. "'Genocide in the 20th Century': The Concept of War and Genocidal Impulses in the Ottoman Empire, 1821–1918." *Holocaust and Genocide Studies* 4, 2 (1989): 175–191.

CON

Whether the perishing of a sizable Armenian population in the Ottoman Empire of late 18th and early 19th centuries constitutes commission of the crime of genocide is a controversial issue of historical scholarship. While a large group of historians and international criminal lawyers rush to call the truly saddening events a genocide, a minority of historians and—quite understandably—Turkish scholars meticulously refrain from any labeling that will require reparation by the Republic of Turkey, which is seen as the successor of the Ottoman state, allegedly responsible for the Armenian genocide.

Those who are unfamiliar with the subject need to recall that there is in fact no denying from even the Turkish side that many Ottoman-subject Armenians died, were killed, or were deported in the late 18th and early 19th centuries. The controversy is more about the nature of the incidents; more precisely, Turks argue that the deaths were simply part of the wartime conditions and had nothing to do with genocide, whereas Armenians are quite sure that the deaths of their relatives was the first genocidal campaign of the 20th century.

Identifying the incidents really matters for both sides—the Armenians and the Turks. For the Armenians, there is no room for further discussion as to whether the deaths were part of a genocide act. This steadfastness and resolve on their stance further determine the very backbone of the sense of being an Armenian. Armenians pay particular attention to the observation of the anniversary of the genocide and ensure its worldwide recognition. On the other hand, Turks feel that attempts to have the allegations recognized as genocide by the international community are part of a greater conspiracy that will undermine Turkey's national interests in the global stage.

That perfectly explains why it is not surprising that the issue of whether the deaths of Armenians in the early 19th century is the outcome of a killing campaign organized and executed by the Ottoman authorities or simply part of conspiracy or plot to put the Turkish state into a difficult position attracts a great deal of attention by historians and international lawyers. This is the exact reason for why similar events such as the French killing campaigns in Algeria or rapid elimination of the native population in North America have gone untouched with special reference to the crime of genocide while disappearance of a large Armenian population became subject of an ongoing debate that focuses on this particular crime.

As noted earlier, most historians and international criminal lawyers tend to see the events under review as genocide. Some even attribute Hitler's eagerness to exterminate the Jews during World War II to the reluctance of the international community to make the Ottomans pay for the genocide they committed against the Armenians, implying that this reluctance served as an incentive. It is tempting to jump to such a conclusion; and unfortunately, this is what this group of historians and lawyers did without inquiring into what genocide really means and whether it is possible to take these events under this heading.

As opposed to the majority group, I firmly believe that Armenians who lived in the Ottoman territories during 18th and 19th centuries suffered a lot and that many were unjustly killed, deported, and subjected to inhumane treatments of various sorts but that we cannot speak of an Armenian genocide for at least three reasons.

Above all, the notion of genocide is not part of our daily language. For this reason, its usage requires certain prior identifications and determinations. To this end, it differs from commonly used notions like massacre, extermination, annihilation, or the like. One may freely use the latter, while use of the term genocide is subject to certain restrictions. Second, commission of the crime of genocide can only be confirmed by a competent international court. This is the greatest shortcoming of the alleged Armenian genocide. Third, historical data are still inadequate to prove that the deaths were genocide. Even if these data seem to be convincing, they need to be verified and taken as the basis for a judgment by a competent judicial body before proceeding with describing the deaths as genocide because of the substantiated genocidal acts.

A thorough and integrated review of historical data will show that the deaths were actually part of intercommunal conflicts, which were common in other parts of the war zones during World War I. Therefore, pro-genocide historians are actually wrong because of their limited focus on the deaths of Armenians while ignoring deaths of the Turks in the same period of time. An integrated approach that will take the latter into consideration will reveal that hundreds of thousands of Turks were also killed by Armenians during the period where a large number of Armenians died of famine, diseases, revengeful acts, or plain killings.

The G-Word and Its Inappropriateness for Frequent Use in Daily Language

The term *genocide* was first coined by Raphael Lemkin (1944), a Polish Jew who spent tireless efforts to ensure that the Holocaust by the Nazis during World War II would be recognized by the international community in a different category of crimes. That is to say, it is not a usual or regular word that has followed the etymological path of any other words extensively used in daily languages. Unlike others we frequently use in our daily lives, the term genocide

connotes an exclusive situation or set of incidents that require certain substantiations and verifications.

Mostly for this reason, its usage is usually discouraged in daily language, and only experts in the relevant fields are viewed as partially competent or authorized for its use. Particularly the media, the state authorities, and scholars pay particular attention to whether it is appropriate to define a certain incident as genocide, considering the repercussions of such an action. In recognition of this delicacy and sensitivity, many are prone to refer to the notion of genocide as "G-word," which recollects that this is such a strong term that it requires prior meticulousness and attention.

Carefulness—and some may even call it timidity—in use of the G-word is visible in most controversial and contentious cases. For instance, the ongoing conflict in Darfur, a war-torn region in Sudan, is not recognized as genocide by the international community because of the strength of the notion and serious repercussions involved in recognizing the incidents as genocide. It should be recalled that even the United Nations Special Commission set up to investigate the international crimes committed in the region concluded that although most heinous crimes have been committed against civilians, the acts by both parties did not constitute commission of the crime of genocide. This is particularly important because the head of the commission, Antonio Cassese, a renowned scholar in the field of international criminal law, was criticized for this conclusion; but interestingly, he does not hesitate to recognize an Armenian genocide committed by the Ottoman Turks. What is not understandable here is Cassese paid particular attention in not using the G-word in regard to such a grave situation like the one in Darfur whereas he almost assumes that the Armenian genocide did actually happen without paying the same attention and utmost care to the historical evidence.

This is in fact the general attitude of those who tend to see the deaths of Armenians in early 19th century in the Ottoman territories as genocide. Despite the restraint that is strongly needed before using the G-word, scholars and politicians, who argue that there was an Armenian genocide committed by the Ottomans, are not so careful. Particularly, the parliaments that proceeded with recognition of the alleged Armenian genocide did not consider the exclusive meaning of the term and the subsequent repercussions of that recognition.

Lack of prior inquiry in the recognition of the alleged Armenian genocide at a number of parliaments as to whether the deaths were really an outcome of a genocidal campaign sponsored by the Turks shows that this issue bears a political dimension. It appears that countries with a substantial Armenian population are more prone to recognize the alleged genocide, whereas states that have good ties with Turkey refrain from any actions or policies that will upset Turkish authorities.

This is most blatantly visible in the United States, where politicians with no administrative positions pledge to spend tireless efforts to ensure official

recognition of Armenian genocide, whereas political figures with influential administrative positions act more rigorously in an attempt to keep relations with Turkey smooth. Especially local administrations—specifically states with large Armenian populations—proceed with swift recognition, while the central government considers the strategic importance of Turkey and its probable role in preservation of American interests in the Middle East.

This suggests that both recognition and lack of recognition of the so-called Armenian genocide are based on a certain political preference rather than legal inquiries or a thorough review of historical facts and evidences. The question here should be to what extent it is wise to use the G-word for achievement of some political goals.

Crime of Genocide and International Law

The same word can have different meanings in daily language and in legal literature. From a criminal law perspective, conviction is possible only if the relevant action is properly defined as a crime under a certain legal system. To this end, it is possible to use a certain word in daily language without considering what it actually means; however, the very same word may fall into a category of crime only if it meets the criteria set to regard it as a punishable offense. For instance, one may call a certain act a massacre, but that does not necessarily mean that the perpetrator of this act will get conviction on massacre charges.

Yet this does not discourage us from using such words or notions indiscriminately simply because they are part of daily language and are used frequently. But the word or notion of genocide is quite different. It is not a word of daily language nor is it appropriate for frequent or common use, as it describes a particular and extreme criminal act.

This brief explanation and clarification should suffice to suggest that in describing a set of actions as the crime of genocide, one should remain extremely cautious and consider whether these actions constitute this crime under proper international legal instruments. Description based on daily usage is not possible considering that the G-word is not one that can usually be used freely by all in different and often uncertain meanings.

To put it differently, the word genocide is only meaningful when it is considered from the international law perspective. This means that it cannot be used freely to describe any particular situation without final judgment by an international legal body. Because it is a crime that is clearly defined under international law, and only subjects convicted of a crime by a competent judicial body are considered to have committed the relevant crime, without a proper international judicial decision, deaths of Armenians may not be defined as genocide.

This implies that in order for a particular situation to fall into the category of the crime of genocide, the relevant acts need to meet the requirements and

criteria set out in an authoritative definition developed through a competent international legal mechanism. Such a definition is provided by the 1948 Convention on the Prevention and Punishment of the Crime of Genocide. The definition was subsequently borrowed by the statutes of the international criminal tribunals, created to deal with the international crimes committed in Rwanda and former Yugoslavia, and the Rome statute, establishing the first permanent international criminal court.

The said convention provides:
[G]enocide means any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such:

(a) Killing members of the group;

(b) Causing serious bodily or mental harm to members of the group;

(c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part;

(d) Imposing measures intended to prevent births within the group;

(e) Forcibly transferring children of the group to another group. (United Nations 1948)

What distinguishes the crime of genocide from other types of crimes and offenses and places it in a special and unique category in this definition is the emphasis on the "intent to destroy." According to the definition, intent to annihilate a group of people should be visible to substantiate the crime of genocide. However, this intent does not necessarily have to be directed against the entire group. The definition states that if there is intent to destroy "a national, ethnical, racial or religious group" "in whole or in part," *and* if the intended action actually takes place, the crime of genocide is considered to have been committed.

In addition to the restrictions in terms of indisputable presence of intent to destroy and commission of the said acts against a certain group, attempts to restrict the size of the group through the acts contained in the definition are not legally regarded as part of genocide. In many cases, acts spelled out in this definition are not considered part of a genocidal campaign because these acts do not seek to "destroy" an existing group, but rather they are committed to restrict the size of the group.

Presence of "intent to destroy" is best visible if there is a clear policy pursued to wipe out a certain group "in whole or in part." This will lead to the conclusion that in order to hold a state responsible for having committed the crime of genocide, presence of a genocidal policy drafted *and* successfully implemented against the targeted group should be legally proven.

In short, substantiation of the crime of genocide requires proof of the presence of two most important and interrelated elements: first, there has to be intent

Armenian refugees, most likely deported from Ottoman Anatolia, at Novorossiisk, Russia, in 1920. (Library of Congress)

to destroy a racial, ethnical, religious, or national group in part or in whole; second, at least one of the acts considered genocidal needs to be actually committed with such intention. I will further discuss whether the deaths of Armenians in the early 19th century in the Ottoman territories fall into the category of genocide with particular reference to these two major elements. What I would like to stress here is that a legal verdict on the commission of genocide has to be released taking the above definition along with the two elements outlined in this definition into account.

Getting back to our particular issue, in order to speak of an Armenian genocide, above all, we need to have a judicial decision taken by a competent international legal body holding that the deaths were genocide. Such a decision also needs to consider these two basic elements of the crime of genocide. In other words, an international legal body should first prove that the Ottoman state had pursued a particular policy to destroy the Armenian population in whole or in part and also actually committed genocidal acts to achieve this goal.

The question that stands here is this: Is there such a decision taken by a competent international legal body? The answer is simply "no." Without such a decision, it is not proper to speak of an Armenian genocide because, as I pointed out earlier, this is not a term usable in daily language. The word genocide refers to a particular and exclusive type of crime, and its presence is proven through a legal verdict alone.

The most important—and one may say the easiest—way of adjudicating the genocide claims is to hear the claims before the International Court of Justice

(ICJ) if allegations are directed against a state. Article 9 of the Genocide Convention of 1948 provides:

> Disputes between the Contracting Parties relating to the interpretation, application or fulfilment of the present Convention, including those relating to the responsibility of a State for genocide or for any of the other acts enumerated in article III, shall be submitted to the International Court of Justice at the request of any of the parties to the dispute. (United Nations 1948)

Therefore, it appears that requesting the ICJ adjudication will undisputedly resolve the issue of whether the deaths of Armenians on Ottoman soil in the late 19th and early 20th centuries was genocide from at least an international law perspective. But until such a resolution has been achieved, claims suggesting that the deaths were genocide obviously are still pending.

But if individuals are accused of having committed genocide, setting up an international criminal tribunal or court vested with the authority to prosecute international crimes will be the most appropriate means to deal with such a situation. In our case, no such option was employed. Therefore, no decision has been made by either the ICJ or a competent international criminal tribunal confirming that there was an Armenian genocide committed by the Turks in the 19th and 20th centuries.

It should be noted, however, that there appeared to be a possibility of establishing an international criminal tribunal to prosecute war crimes in the Paris Peace Conference of 1919 convened by the victorious powers of the war. At the conference, a commission was set up to address the crimes committed during the war. It concluded that individuals should be held responsible for violations of the laws and customs of war and of the laws of humanity, with an additional view that an international criminal court should be created in the cases where the accusations against the suspects were directed at more than one state.

At the conference, the victorious powers also discussed Germany's surrender and negotiated a treaty where they dictated the terms. The deliberations at the conference also included the issues concerning the prosecution of high-level officials and war criminals from the defeated powers for "crimes against the laws of humanity." At the end, the Treaty of Versailles was concluded with Germany on June 1919. The treaty provided establishment of an ad hoc international criminal tribunal to prosecute Wilhelm II of Germany for initiating the war. However, Kaiser fled to neutral Holland, which refused his extradition.

The Treaty of Versailles also created the Commission on the Responsibilities of the Authors of War and on Enforcement of Penalties to investigate the responsibility for initiating the war and the violators of the laws of war. The crimes considered by the commission included rape, use of poisonous gas, murders, massacres, and waging aggressive war, for which it proposed a tribunal consisting of 22 members. The commission completed its work and submitted its report, which consisted of a list of 895 alleged war criminals.

A similar path was followed through the Treaty of Sevres, which was never enforced. This brief explanation suggests that although the Allied Powers considered prosecution of international crimes committed in World War I, they never took this issue seriously, missing a great opportunity that would put an end to current discussions on the alleged Armenian genocide.

But I would like to stress that accusations directed against Germany in regard to the commission of international crimes in World War I were far more serious than those directed against the Ottoman Turks. For instance, in a letter addressed to the Austrian Kaiser Wilhelm II, Kaiser of Germany, wrote:

> My soul is torn, but everything must be put to fire and sword: men, women and children and old men must be slaughtered and not a tree or house be left standing. With these methods of terrorism, which are alone capable of affecting a people as degenerate as the French, the war will be over in two months, whereas if I admit considerations of humanity it will be prolonged for years. In spite of my repugnance I have therefore been obliged to choose the former system. ("What Punishment Fits" 1919)

Despite this, so-called Armenian genocide becomes popular in the present time because of political reasons, whereas no reference is made to German atrocities in the same period. This undoubtedly proves that the Armenian genocide claims are promoted for political reasons, which are taken into considerations by national parliaments proceeding with recognition of an Armenian genocide.

Do Historical Facts Really Support Armenian Genocide Claims?

Those who allege that Ottoman Turks committed crimes of genocide against the Armenians in early 20th century believe that a verdict by an international tribunal is not needed to confirm their allegations simply because there are plenty of historical evidences on this matter. Mostly for this reason, promoters of the alleged Armenian genocide pay particular attention to keep the issue away from adjudication of an international court.

But it should be noted that their overreliance on historical findings to make their point is also groundless because while they consider the facts or evidences supporting their claims, they fail to take the stories of the other side into account. In other words, assessing the whole story from the perspective of murdered Armenians will simply be misleading because there are so many Turks killed by Armenians in various parts of the Empire; but their stories were ignored by the supporters of the Armenian genocide claims.

More important, historical findings and facts should be rechecked for the purpose of determining whether these would suffice to call the deaths of Armenians in the early 20th century genocide. Two particular and most important

criteria of genocide should be taken into consideration. Otherwise it would not be appropriate to describe the deaths as genocide just because many Armenians were slaughtered or massacred.

Central to reaching a decision as to whether a set of actions constitutes the crime of genocide is whether the relevant actions have taken place under a pre-set and well-defined plan drafted by the state or state authorities. To this end, in order to speak of the presence of genocide, there should be a state policy pursued for a specific purpose of destroying a certain religion, national, racial, or ethnic group *and* this policy should be actually implemented and result in whole or partial destruction of the targeted group.

In our case, there is no indisputable evidence suggesting that the Ottoman state or statesmen had drafted a well-crafted plan to annihilate the Armenians in the territory of Ottoman Empire. More important, there is also no evidence indicating that such a policy—if ever drafted—had been implemented to wipe out the Armenian population.

Supporters of the Armenian genocide claims often cite the deportation decision made by the Ottoman authorities as an indication of their coercive decisiveness to send them to this deadly journey. They assert that deportation or displacement of such a large population would mean nothing but death for hundreds of thousands people because of the conditions prevalent at that time.

This explanation seems a bit logical and reasonable given that most of the casualties were attributed to this journey, which became a deadly walk for many Armenians. But the Ottoman state or the state authorities should not be accused of devising a plan of genocide to exterminate the Armenian population just because they made such a decision. This might be a terribly wrong decision; but this does not necessarily mean that it could be taken as clear evidence for the existence of an intention to destroy Armenians in whole or in part.

This was a decision that the state authorities of the time considered would be the best solution to the ongoing problems in the regions with a substantial size of Armenian population. Many Armenians suffered because of this decision; but that does not indicate liability of the state authorities for preparing a great plot to eliminate all Armenians.

Those who allege that Armenians were subjected to a genocidal campaign by the Turks also make reference to the trials and conviction of several Ottoman officers, who were executed because of crimes committed against Armenians, at Ottoman courts. True, some military and administrative officers were tried and executed in connection with offenses against civilian Armenians; but this does not suggest that they were responsible for commission of the crime of genocide. There is a broad set of war crimes, none of which may be considered as genocide; therefore, these trials might be relevant to such offenses.

More important, because such trials were held by the Ottoman state, this could actually prove that there was no centrally devised plan to annihilate the

entire Armenian population. If there had been such a plan, the state could not have held its own officers responsible for these offenses.

It should also be noted that the Armenian people and the Turks have coexisted peacefully for centuries in the Ottoman territories. That is to say, there has never been a culture of hatred or enmity held by Turks against the Armenians. Even though there was some hidden or indirect hostility between the Greeks and the Jews, most subidentities were tolerant of one another. Among these subidentities, Armenians were the closest to the Turks in terms of cultural and lifestyle similarities. A foreigner would have difficulty discerning an Armenian from a Turk even as late as the 19th century.

In return, Armenians have never considered rioting against the administration, mostly held by Turks. For this reason, they were referred to as the "loyal nation" (*Millet-i Sadıka*). Considering this loyalty, late Ottoman sultans appointed some high-level Armenians to influential posts. Even Abdulhamid II, who is abhorred by Armenian diaspora, had affinity with Armenians and reserved some crucial positions for leading Armenian figures. Despite an assassination attempt plotted by Armenian terrorists against him, Abdulhamid II has never considered wiping out the entire Armenian population, nor did he remove influential Armenian figures from government posts.

Of course this cannot be taken as an assurance that Turks would never commit the crime of genocide against Armenians. But this very crucial fact implies that historians should consider the particular nature of relations between Turkish and Armenian communities in the Ottoman Empire when making generalizations about the Armenian genocide claims. That is to say, if there has been no source of conflict or tension for centuries between these communities, then there should be no particular reason for the breakout of a deadly conflict that left millions of dead people.

At this point, recalling the tensions between conflicting communities in two most important and legally confirmed genocides will be illustrative. In regard to the Holocaust, it could be said that the Jews had suffered from certain stereotypes held by a substantial group of German people and that there had been no peace between these two communities. Hitler successfully mobilized the anti-Semitic sentiments to ignite a genocidal campaign against the Jewish population. Likewise, in Rwanda, Hutu extremists exploited the long-standing tension with the Tutsi minority and justified an extremely violent campaign by which they aimed to wipe out the Tutsi identity. Extremists made particular references to the alleged abuse by Tutsis, whose identity was formerly promoted by colonial Belgians.

But in the case of Armenian deaths, there was no such motive that could be used by the Turks and the Ottoman authorities. What happened was: a conflict broke out between the parties because of the war conditions and the subsequent deportation of a large Armenian population, which resulted in the deaths of

hundreds of thousands of Armenians. It would be explanatory to recall once more that an almost equal number of Turks had been slaughtered by Armenians in their hopes of having an independent state of their own, encouraged by the Russians, who occupied a substantial part of the eastern Ottoman territories.

Most historians are prone to focus on what happened to the Armenians during early the 20th century because their minds are set on investigating the facts about the alleged Armenian genocide. This is a legitimate yet insufficient inquiry because it misses the details of the other side of the whole story. A brief survey of what happened to the Turks in the same period of time and of crimes committed by the Armenians against the Muslim population will prove that the killings were not genocide but rather a wholesale conflict that left many Muslims and Armenians dead in the warzone. The fact that nothing harmful was done to the Armenians living in other parts of the Ottoman state confirms this, as there was no such conflict in Istanbul or other provinces in Marmara or the Ege region.

Therefore, an integrated historical approach should consider the whole picture to have a more accurate depiction of what happened. Supporters of Armenian genocide claims often fail to take the war conditions into account and the fact that the Ottoman state lost authority in eastern and southeastern parts of the country, where most of the massacres took place. It should be once more noted that these atrocities were committed by both sides; considering that the death toll was growing and proved not to stop growing any time soon, the Ottoman state decided—for better or worse—to relocate the Armenian population in an attempt to put an end to the clashes.

The process of relocation was unfortunately deadly, with countless sad stories of not only famine and excessive heat, but also of revenge attacks by armed Kurds, who chased the deportees and caught many Armenians defenseless on their way to another place of habitation.

Conclusion

The issue of Armenian genocide remains one of the most controversial historical debates because of its political dimensions. Since genocide charges are not acceptable by the Turkish government and the Armenians are prone to base their identities on the genocide claims, there is no apparent prospect for a lasting resolution to the issue to which both parties can agree.

That being said, the genocide allegations are not well grounded because of at least three fundamental reasons. Above all, because it is a rather technical and legal term, genocide is not appropriate for use in daily language. In other words, one cannot use this word freely to define a certain set of actions as genocide because it requires a legal verdict by a competent international legal body. Second, without such a decision, whether the relevant actions are genocide will remain pending. In our case, there is no such decision confirming that the Armenian

---

New York Life Settles the Armenian Claims

After a long and grueling legal battle, on January 27, 2004, a "Settlement Agreement" was signed by the New York Life Insurance Company with descendants of policy holders who had been killed in Armenia in 1915.

Overall some 8,000 to 9,000 Armenians living in the Ottoman Empire had purchased life insurance policies, and after the killings in 1915, the New York Life Insurance Company used a local attorney to track down policyholders and their families. Payments were made to about a third of the policyholders. However, there were problems with many other payments because of the inability of some of them to furnish death certificates and other paperwork.

The 2004 settlement involved the payment of $11 million to the heirs of 2,400 policyholders, $4 million in costs to the four law firms who brought the case, $2 million to cover the costs of the advertising of the case and the hiring of agents to track down heirs, and $3 million to a range of Armenian charitable organizations.

---

population was subjected to a genocidal campaign at the hands of Ottomans. Third, historical research often focuses on one side of the whole story to see whether the deaths were genocide. A thorough investigation inclusive of massacres committed by Armenians against Muslims will show that what happened in the early 20th century in the Ottoman territories were brutal killings, but not genocide.

References and Further Reading

Charny, Israel W., ed. *Encyclopedia of Genocide*. Santa Barbara, CA: ABC-CLIO, 1999.

Fein, H. *Genocide: A Sociological Perspective* Thousand Oaks, CA: Sage, 1993.

Kuper, L. *Genocide: Its Political Use in the Twentieth Century*. New Haven, CT: Yale University Press, 1981.

Lemkin, Raphael. *Axis Rule in Occupied Europe*. Washington, DC: Carnegie Endowment for Peace, 1944.

United Nations. "Convention on the Prevention and Punishment of the Crime of Genocide." December 9, 1948. Available at www.hrweb.org/legal/genocide.html (accessed June 18, 2010).

Vohrah, Lal Chand, et al. *Man's Inhumanity to Man: Essays on International Law in Honour of Antonio Cassese*. The Hague: Kluwer Law International, 2003.

"What Punishment Fits the Ex-Kaiser's Crime?" *Literary Digest* 61, 4 (April 26, 1919).

5

The British had shipped weapons aboard the *Lusitania*, in effect using women and children as "human shields" for a war cargo.

PRO   Ardhana Mudambi
CON   Justin Corfield

PRO

Both in Ireland and in the United States, the sinking of the *Lusitania* served propagandists' purposes to convince constituents of both nationalities to support Great Britain against Germany in the Great War. At the outbreak of the war, Irish nationalists naturally supported the kaiser against the common enemy. The Crown, therefore, trying to undermine these sentiments, had, even before the sinking, vilified German U-boat crewmembers as savages who set out for rape and pillage (Molony 2004: 12). With the tragedy came the opportunity to prove their claims by vividly painting a mass murder scene that had occurred only 12 miles off the shore of Ireland. Leaflets called Irishmen to send out soldiers in the names of the women and children who had died so close to home due to German depravity.

Although the United States held an ostensible position of neutrality, the sinking rapidly changed American attitudes as well. The *New York Times* quickly depicted the sinking as a wanton massacre of American civilians ("Sinking Justified" 1915). A two-page spread featured prominent Americans who had lost their lives in the sinking. Subsequently, another two pages featured the burials of these victims. Americans, some of whom had previously sympathized with the Germans, now detested the kaiser for his uncivilized war tactics. The death of innocent women and children could not be acceptable in Western warfare (Schmidt 2005: 75).

On May 7, 1915, off the coast of Kinsale, Ireland, 1,201 men, women, and children of the 1,959 passengers lost their lives when U-20, a German submarine, sank the *Lusitania*, a supposed British cruise liner traveling from New York to Liverpool, England. Among the victims were 128 Americans. In sinking the cruise liner, Germany had ostensibly acted against International Cruiser Rules, which allowed nations to destroy enemy ships only when carrying contraband, defined as "munitions, that is weapons or goods that could directly aid an enemy war effort" (Schmidt 2005: 75). Cruiser Rules also mandated that such destruction

107

should occur only after allowing the crew and passengers to disembark. The *Lusitani*a was ostensibly a civilian ship carrying civilian passengers, including those from a neutral country and, therefore, should not have been sunk in the first place. Hence, the sinking was a heinous German act of willful murder, or was it?

According to German officials, the sinking of the *Lusitania* was indeed a heinous act of wanton murder, but not a German one. Accordingly, the *Lusitania* carried contraband and was under the authority of the British Admiralty, thereby giving Germany the right to destroy the cruise liner without regard to the passengers. Germany further accused the Crown of using Americans as a "cloak," thereby transferring the sin of murder to the British. U.S. Secretary of State William Jenkins Bryan echoed German sentiments. Bryan condemned the British for "relying on passengers, including men, women, and children, of the United States to give immunity to a vessel carrying munitions of war." Bryan defended German rights to prevent contraband from reaching the Allies and paralleled British use of the *Lusitania* to transport contraband to the placing of "women and children in front of an army" (quoted from Schmidt 2005: 80). In effect, the secretary of state accused the Crown of having used women and children as shields for its munitions.

In response, the Crown conducted an investigation of the sinking, concluding that "the whole blame for the cruel destruction of life in this catastrophe must rest solely with those who plotted and with those who committed the crime" (Molony 2004: 184), thereby, denying charges that Great Britain had carried contraband upon the ship and placing guilt squarely upon the kaiser's shoulders. Subsequent Irish and American swings favoring the British demonstrate that world opinion sided with the Crown.

However, this section's discussion goes against common perception, arguing that the British had shipped weapons upon the *Lusitania*, in effect using women and children as human shields for war cargo. First, the section demonstrates that the British investigation must be disregarded due to its lack of substance. Then, the section reviews *Lusitania*'s original purpose as a warship for the admiralty. Next, it demonstrates the ship's further metamorphosis in order to carry contraband during World War I. Finally, it presents proof of armaments upon the ship.

British Investigations

Although British inquiries to determine whom to blame for *Lusitania*'s destruction demonstrated that the British had not violated international laws (i.e., had not carried contraband upon the ship), British investigations cannot be trusted. First, the British could not afford negative publicity and, therefore, had no choice but to hide unfavorable truths. Furthermore, the actual investigation demonstrates a superficial treatment of the sinking. Finally, the Crown has kept various related documents out of public reach.

First, British inquiries cannot be trusted because the loss of moral support of its own people, the Irish, or especially the Americans, could mean losing the war. Although President Woodrow Wilson claimed neutrality, in actuality, he lent great support to his Anglo brethren. By 1917, U.S. exports to Germany had been virtually eliminated. From 12 percent of total exports in 1914, the total percentage of exports to Germany had dropped to less than 1 percent by 1917. On the other hand, exports to Great Britain had increased from 25–33 percent of total exports, demonstrating that Wilson's neutrality was nothing more than lip service. In spite of America's aid, by *Lusitania*'s sinking, the British was down to six weeks of food and was in dire need for armament (Schmidt 2005: 78–79). Undoubtedly, the Crown could not afford U.S. support turning to Germany, and after *Lusitania*'s sinking, Americans were out for blood.

Further demonstrating that British investigations could not be trusted, the actual inquiries exhibit a superficial treatment of the sinking. Only three years earlier, after the *Titanic*'s sinking, British inquiries spanned 36 days with more than 25,000 questions asked to pertinent individuals. In contrast, the investigation covering the *Lusitania* encompassed a paltry 103 pages (including title pages and an index of witnesses), a result of only 2,312 questions asked (Sauder 2005: 105). If indeed the British were innocent victims, why wasn't a more thorough investigation conducted akin to that of the *Titanic*?

Finally, British investigations are unreliable because the government successfully made key documents related to the sinking disappear. For instance, when Admiral Graham Greene asked about the ammunition carried by the *Lusitania*, Oswyn Murray of the admiralty responded "Curiously enough I had some difficulty in laying my hand on any document containing the exact facts" (quoted from Preston 2002: 384). Hence, information concerning ammunition was hidden.

Of course, this argument brings us to the question of present day. Have these missing documents been released? Can a conclusive statement be made regarding the issue of contraband by researching these papers? Unfortunately, the Crown continues to keep the evidence under lock and key. While some of the previously missing papers were purchased by the Cunard archives, official files continue to be fleeting; one moment, they give leads to understanding the tragedy better, then they disappear. Blank sheets inserted in the official dossier give reason to suspect missing documents such as telegrams. Researchers doubt even the authenticity of newly released files. Hence, the British government can obviously not be trusted to give accurate information about the *Lusitania*'s destruction, and, therefore, her investigations are not to be trusted.

Churchill and the Sinking

An additional argument against the idea that women and children were used as human shields for war cargo stems from the idea that the Crown purposely meant

for the *Lusitania* to be attacked in order to draw the United States into the war. This theory of course, then worsens the blame placed on the British government since women and children now have been relegated from human shields to human sacrifices; this argument, in turn, however, supports the theory that the sinking was in fact the fault of the British. Although this section does in fact argue that the British government placed women and children in harm's way, it does not go so far as to claim that the government ensured their murder. However, a glance at the argument is worth the reader's while.

If indeed, the *Lusitania* carried important war cargo, why was the ship not further protected? An escort vessel was not provided. Furthermore, evidence demonstrates a paltry supply of information to Captain Turner, the ship's leader. If the admiralty had provided him with more information, the *Lusitania* might have avoided destruction. These questions suggest that the Crown may have purposely condemned the *Lusitania* to her death.

First, the *Lusitania*, like other similar ships, was expected to fend for herself using her immense speed. Even after her sinking, the admiralty did not provide further protection to other similar ships, citing the importance of speed as their first line of defense. This tactic continued as well through World War II. Furthermore, international laws dictated that passenger ships could not be destroyed without a stop and search for contraband, an issue discussed shortly. By providing naval support, the *Lusitania* would lose this privilege (Preston 2002: 399).
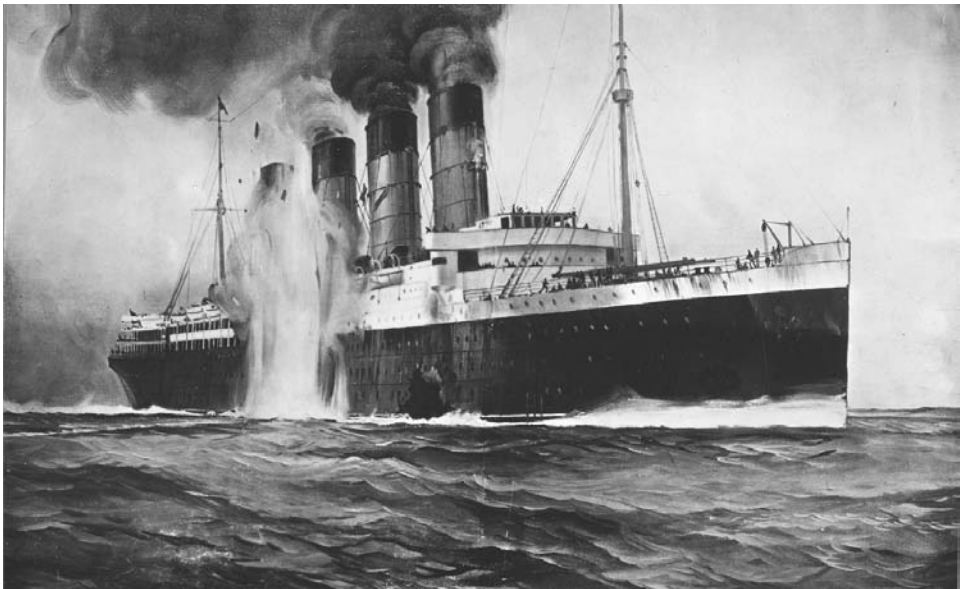
The Cunard passenger liner *Lusitania* was torpedoed and sunk by a German U-boat on May 7, 1915, off the south coast of Ireland, en route from New York to Liverpool, with the loss of 1,201 lives, 128 of them American. (Library of Congress)

Furthermore, historians generally concur that little information was presented to Captain Turner, leaving him at a loss in his struggle against the German submarine U-20. Why did the British government not make a greater attempt to forewarn Turner, especially if the *Lusitania* contained necessary war cargo? Overall, this lack of action resulted from a failure in leadership. The First Lord of the Admiralty, Winston Churchill, liked to know everything transpiring and to make decisions alone. During *Lusitania*'s travel, he was more occupied with embroiling Italy in the war as an ally than with the ship. The second in command, Admiral Fisher, by admission, was near a nervous breakdown and could not at the moment handle the pressures of the *Lusitania*. Other officials, beneath the two, were far too scared of Churchill to act alone; hence, the upshot was a vulnerable *Lusitania*.

The bulwark of the Lusitanian conspiracy theory is none other than Churchill, at the time a young nobleman with grand political ambitions and very slowly developing prudence. On February 12, 1915, he commented to Walter Ranchman, president of the Board of Trade, that "it is most important to attract neutral shipping to our shores in the hope especially of embroiling the U.S.A. with Germany" (quoted from Preston 2002: 395). Churchill could undoubtedly have been ruthless enough to perform such a horror. He was one of the first to support bombing military targets within populated areas, using poison gas against enemies, and violating Dutch and Danish neutrality in order to enter Germany. He even suggested sacrificing loyalty to his long ally Portugal by presenting it as a bribe to bring Spain on to the side of the Allies. Most condemning, after *Lusitania*'s sinking, Churchill believed the tragedy had in fact worked in Britain's best interests.

Nevertheless, wouldn't Churchill have, after a more careful inspection, opted for other, more promising venues to defeat the Germans? After all, embroiling the United States in the war would have actually given the Germans the upper hand. As previously stated, by 1917, the British were in dire need of food and armament, and one-third of America's total exports were in fact to Great Britain. If Lady Liberty had entered the war immediately (which fortunately for Great Britain, she did not), the new participator would in fact have wanted to keep her ammunition for her own use, thus opening the possibility for Germany to triumphantly invade a helpless Great Britain (Schmidt 2005: 80). While holding true to Churchill's eventual belief that the sinking of the *Lusitania* actually aided Great Britain in terms of propaganda, the upshot of the sinking could not have been predicted. Hence, to believe that *Lusitania*'s torpedo was in fact the result of British conspiracy is more a topic left for the imagination and film screen than for a historical argument.

Creation and Metamorphosis of the *Lusitania*

The creation and later metamorphosis of the *Lusitania* in themselves strongly suggest that the ship carried munitions. First, the liner was built only after the government lent a large sum of money to the Cunard Steamship Company in

exchange for future naval use. Furthermore, during the war, passenger sections of the liner were gutted by the admiralty to create additional cargo space. Hence, the *Lusitania* had a clear naval purpose.

Suggesting that the *Lusitania* did in fact have ammunition on board, the liner was built under government subsidy for naval use during war. Why would the government spend money on a ship and then not use it during war? British enthusiasm for the construction of the *Lusitania* was partly to show up the much younger German Empire since in 1897, the latter with its construction of Kaiser Wilhelm der Crosse, had in fact challenged the Crown's position as the ruler of the seas. As the Crown's naval prowess was a sine qua non of its claim as the world's paramount power, it could not allow another empire to demonstrate itself stronger in the waters.

Hence, the British government made an agreement with Cunard Steamship Company's chairman Lord Inverelyde in order to build two new cruise liners that could, during war, provide the trump cards the British needed to win. It offered the company a loan of 2,600,000 pounds at 2.75 percent interest for the construction of the *Lusitania* and *Mauretania*. The Crown also offered an additional annual subsidy of 150,000 pounds to maintain the ships ready for war. In turn, the Cunard Steamship Company had to agree to certain conditions:

1. Cunard had to remain a wholly British company for the next 20 years, thus, prohibiting any other nation from benefiting during war from Britain's investment.
2. The British government, at will, could appropriate the two liners during war.
3. The vessels must be equipped to fit 12, six-inch, quick-fitting guns.
4. Coal bunkers were to line the ships in order to absorb enemy gun fire.
5. Not less than 50 percent of the crew had to belong to the Royal Navy Reserve or the Royal Navy.

Given the money and prestige connected to the *Lusitania* and *Mauretania*, how likely would it be that the British would not use its auxiliary cruisers in event of war? In fact, it did very openly use the *Mauretania*, first as a trooper and subsequently a hospital ship. But what about the *Lusitania*? Was the British government going to allow their invested money to sink by allowing the liner to serve only passengers? Of course not (Sauder 2005: 10).

Furthermore, the *Lusitania* was repeatedly listed as a Royal Navy ship. One could find the *Lusitania* in *Jane's Fighting Ships* for 1914. Furthermore, in *Brassey's Naval Annual 1914,* the *Lusitania* could again be found as a "Royal Naval Reserved Merchant Cruiser" (Preston 2002: 386). Of course, the key word here is *reserved*, meaning that the *Lusitania* might not be used for war purposes, but its subsequent metamorphosis demonstrates otherwise.

The Admiralty's Trade Division decided to use the ship as an express cargo carrier for priority government supplies. The admiralty transformed the majority of the third-class accommodation in the lower forward section of the ship into a warehouse for additional cargo. This metamorphosis significantly aggrandized the forward cargo hold. Other sections of the ship as well were placed on hold for the admiralty's use; the British government specially insured this special, noncommercial space. The remaining room was of course used for passenger accommodation, permitted on *eastbound* trips (Peeke 2002: 43). Furthermore, the ship's captain received direct navigational orders from the senior naval officer in New York or Liverpool preceding every departure, hence placing the ship under the admiralty as the Germans had claimed when defending their right to sink the ship.

Finally, although the Cunard Steamship Company was losing nearly 2,000 pounds on every trip, the admiralty would not allow the *Lusitania* to abstain from its scheduled runs. What did the *Lusitania* carry that made its run so important? What were in fact the "priority government supplies"? If the cargo was anything but food, the British had in fact violated Cruiser Rules, and common sense does in fact dictate that the supplies were more than just food. Hence, the British did in fact use women and children as human shields for their war cargo.

Proof of Munitions

As aforementioned, the British by 1917 were in dire need of armament. Within two months of the war, Lord Kitchener alerted senior government ministers that the Redcoats in France were using more ammunition seminally than the British factories could produce. Their arms stocked would last for only three months, and the British were in desperate need for alternate suppliers. Hence, to believe that munitions were not stored in the additional cargo space of the liner would be counter-intuitive. However, history does not always follow intuition.

According to rules set up by the Hague Convention of 1907, by which all Western nations were bound during war, neutral countries were not allowed to supply belligerent nations with any sort of war cargo. However, the United States's neutrality was only skin deep. U.S. banker John Pierpoint Morgan, Jr., craftily set up an exchange between himself and the admiralty. The Admiralty Trade Division (the same division that created the additional cargo space in the *Lusitania*) placed purchase orders for munitions with J. P. Morgan. The banker, in turn, using a facade of nonexistent companies, provided the goods. Ships that the admiralty controlled became the method of transport. Their passengers became sitting ducks for German U-boats.

In order to clear customs, the British filed a false manifest with D. F. Malone, the U.S. Collector of Customs in New York, in which all mentions of ammunition were omitted. Malone, President Wilson, and the new secretary of state Robert

Lansing knew of the prevarication; however, they tacitly approved the operation. To prevent further suspicion, the ship always filed a second manifest after leaving U.S. waters, more specifically outlining cargo. This second manifest could enumerate "machine parts, metallic packages, or possibly furs or extra large quantities of . . . cheese or butter" (Peeke 2002: 45). Machine parts were obviously those used for rifles and guns; large amounts of fur, cheese, or butter were not of paramount importance during a war and were obviously covers for something else. When a cover is required, the actual identity has to be illegal.

In reality, the items proven to be transported on ships such as the *Lusitania* included military goods such as gun cotton, artillery shells, bullets, fuel oil, brass, copper, metallic powders for making ammunition and explosives, fuse mechanisms for shells, vehicle parts, uniforms, motorcycles, airplane parts, and so forth, all illegal war cargo.

Furthermore, the second manifest for the *Lusitania* included weapons and ammunition including 2.4 million rounds of bullets. Upon a dive sponsored by American businessman F. Gregg Bemis Jr., two cases in the bow section of the ship revealed 15,000 rounds of .303 bullets. Rifle cartridges and shrapnel cases were also listed on the manifesto. Hence, even if nothing further was on the ship, the *Lusitania* did carry contraband, and passengers were, therefore, converted into human shields.

Controversy over Munitions

Much of the controversy over the contraband rests with the kind of contraband carried by the *Lusitania*. The argument rests over whether it had guns or if the furs and cheese were really furs and cheese, and whether witnesses were honest or guileful in their recollections. This section looks at these controversies, subsequently arguing that these details are in fact superfluous.

First, the argument exists that the British government was not to blame because the main witness who testified that there were guns aboard the ship was found untruthful. Gustav Stahl testified upon German instigation to the U.S. government that there were guns aboard the ship and later was arrested for perjury. No other passengers recalled seeing mounted guns nor found them elsewhere. A movie made in New York about the *Lusitania*'s departure also did not capture the guns. Finally, customs officials did not recall them (Preston 2002: 386–87). However, if the guns were in fact hidden, most passengers would not have seen them, and a film depiction certainly would not have captured them. Custom officials, as previously discussed, were already in J. P. Morgan's pockets. Hence, to believe their words would have been naive.

Some historians argue that the guns were in fact hidden to be mounted when necessary since the deck was equipped with six-inch gun rings. However, to mount these supposed guns would have taken 20 minutes. Furthermore, there was little need for guns since, as aforementioned, the *Lusitania*'s best and foremost

line of defense was her speed. Hence, most probably *loaded* guns ready for immediate use did not exist on the ship. Nevertheless, this argument does not preclude the transport of unloaded guns for later British use.

Furthermore, some historians argue that the furs and cheese noted to be hiding contraband were actually furs and cheese. For instance, some furs washed up on the Irish coast following the sinking. Furthermore, the German government could never find any conclusive data supporting the accusation of contraband not listed on the manifest being smuggled aboard. However, why would Britain, in the middle of a war in which she was running out of food, want to carry furs home? While a few furs may have truly existed as a decoy, the idea that the British authorities would use up precious cargo space with furs rather than food (which would not have had to be hidden) or armament makes little sense.

Finally, historians such as Diana Preston argue that although contraband was clearly in the ship, as stated in the second manifest, the Germans still did not have the right to torpedo the ship. In accordance with international laws, the Germans should have allowed for the evacuation of the ship and then searched it. After securing the lives of the passengers, they would have, indeed, been within their rights to destroy the ship. However, an honest look at this reasoning quickly demonstrates the holes in this theory.

The British had long been ignoring international rules. They had used neutral flags to confuse the German U-boats. They had created an illegal blockade, preventing food from reaching the European continent. Some 750,000 German civilians, including women and children, died as an upshot of this blockade! The British also arbitrarily changed the list of illegal goods, seizing random items such as cotton and oil, making shipping for neutral countries prohibitively expensive. Furthermore, under Churchill's authority, as the Germans allowed crews to disembark their boats before subsequent destruction, the crews would illegally attack the U-boats by ramming into them or by using hidden weaponry. Captain Turner of the *Lusitania* had in fact those instructions as well (Schmidt 2005: 76–77, 79).

Were the Germans not within their right to protect their own people? Had Churchill's actions not forewarned the kaiser against normal, gentlemanlike behavior? As U.S. Secretary of State Williams Jennings Bryan said upon his resignation, "Germany has a right to prevent contraband from going to the Allies," and undoubtedly, she had a right to do so without fearing British guile.

No doubt, the British rested on beliefs that Germany could not overtly go against established Western traditions of protecting innocent women and children. (The British, of course, had caused the death of many through the blockade, but the direct connection between the British and the deaths of innocent German civilians was less apparent.) Therefore, juxtapositioning women and children with illegal contraband protected the ship, and regardless of whether or not the contraband included loaded guns or undocumented ammunitions rather

---

The *Lusitania* Medal

Much British propaganda after the sinking of the *Lusitania* focused on the "*Lusitania* Medal," which was minted by the Munich metalworker Karl Götz. He made about a hundred of these, which showed the ship with the message "No Contraband" yet the ship's deck clearly showed planes and weaponry, barely leaving enough room for passengers.

Reports of the medal were published in the British press soon afterward with claims that although the ship was shown carrying planes (which were not on board), no mention is made of the civilian passengers who certainly were on board. The German government questioned Götz and found that he had made the medals in an effort to be critical of the British claims that the *Lusitania* was purely a passenger vessel. But the German authorities decided that they would prevent distribution of the medal.

However, although the Germans had prevented its distribution in Germany, the British did get one of them, and Gordon Selfridge, the owner of the department store Selfridges, made some 250,000 copies, which were then sold to increase awareness of the incident and also to raise money for British wartime charities.

---

than furs and cheese, the mere fact that contraband existed meant that women and children had been used as human shields to protect war cargo.

Conclusion

A ship carrying contraband should not rely upon passengers to protect her from attack—it would be like putting women and children in front of an army.

—William Jennings Bryan

How would Americans or the Irish have felt if the British has whisked away innocent civilians to stand in front of the trenches, blocking the bullets? Would they have still argued that the Germans should not have fired? Would historians quibble over the color of their clothes or how well loaded the British guns were (not whether they were loaded, but how well)? As this section demonstrates, arguments over whether or not the British used innocent women and children as shields for a war cargo amounts to analogous mockery. There are only three important facts in question.

1. The *Lusitania* was in fact built as a naval ship.
2. The *Lusitania* did, by confession in the second manifest, carry contraband.
3. Common practice dictated that the Germans would not fire against women and children.

Whether or not there were loaded guns in the ship or the supposed furs and cheese were actually additional contraband hardly alter the aforementioned three facts. Nevertheless, however wrong the British were, their tendentious newspaper articles, brochures, and so forth helped anger the U.S. and Irish public against the Germans. The subsequent British victory helped, in turn, to turn the propaganda into history.

References and Further Reading

Molony, Senan. *Lusitania: An Irish Tragedy*. Douglas Village, Cork, England: Mercier Press, 2004.

Peeke, Mitch, Kevin Walsh-Johnson, and Steven Jones. *The Lusitania Story*. South Yorkshire, England: Leo Cooper, 2002.

Preston, Diana. *Lusitania*. New York: Walker, 2002.

Sauder, Eric. *RMS Lusitania. The Ship & Her Record*. NP: Tempus Publishing, 2005.

Schmidt, Donald E. *The Folly of War*. New York: Algora, 2005.

"Sinking Justified, Says Dr. Dernburg." *New York Times* (May 9, 1915).

CON

Since the sinking of the British Cunard liner *Lusitania* on May 7, 1915, with the loss of lives of 1,201 men, women, and children, there has been much debate over whether or not the British government was deliberately using the passenger liner as a way of shielding the otherwise hazardous operation of bringing war materiel across the Atlantic Ocean. Evidence that the ship was carrying military supplies emerged immediately after the sinking, being published in the *New York Times* in its coverage of the sinking itself. As a result of this, there were some recriminations soon after the sinking of the liner, but the transporting of weapons was soon lost in the other horrors of World War I such as the loss of 19,000 British soldiers on one day of battle just over a year later on the first day of the Battle of the Somme.

It was the publication of Colin Simpson's best-selling book *Lusitania*, in 1972, that led to the accusation that the British government deliberately used the *Lusitania*'s passengers as a human shield becoming an "accepted fact." Indeed the sinking of the liner is also often cited as the major reason for the United States entering World War I, in spite of the fact that the U.S. declaration of war on Germany did not come until April 6, 1917, nearly two years after the *Lusitania* was sunk, and there were many other reasons for the eventual U.S. declaration of war, although British propaganda did make much of the sinking of the vessel and the large number of U.S. civilians who died as a result.
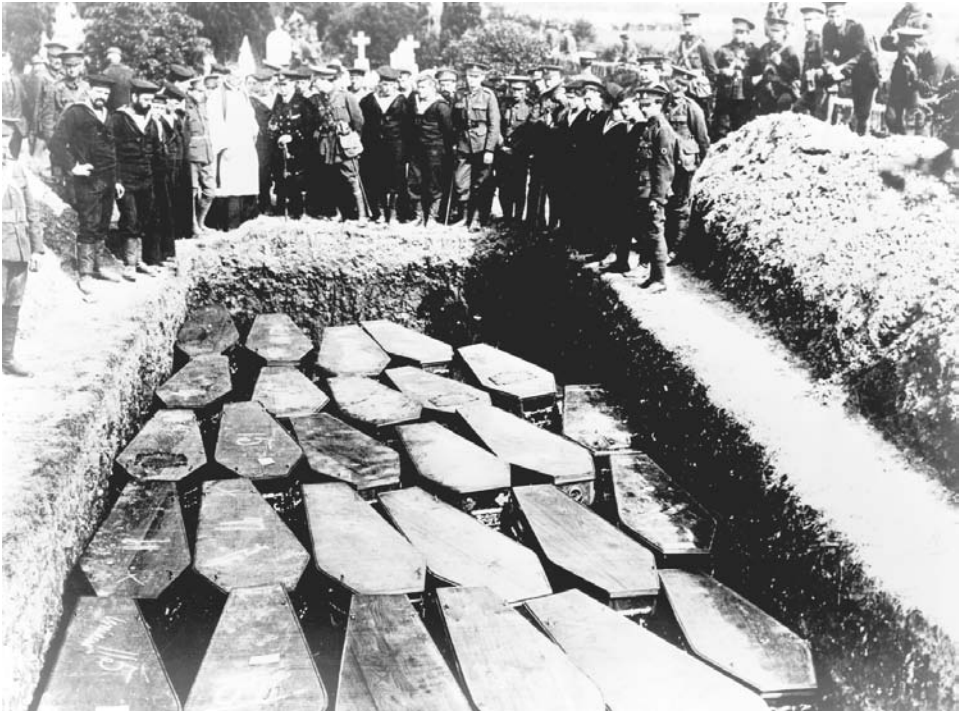
Victims of the *Lusitania* sinking, after recovery from sea, are buried in Ireland in a common grave. The *Lusitania*, a passenger ship, was sunk by torpedoes from German U-boats in May 1915. (Library of Congress)

Rather than a deliberate use of the passengers on the *Lusitania* as a human shield, transporting war materiel across the Atlantic was so important to the British war effort that callous disregard of the passengers seems far more likely. Indeed the plans to use the *Lusitania* for transporting war supplies goes back to its original construction, as well as that of many other large vessels, with the cost being subsidized by the British government.

The *Lusitania* itself was built at Clydebank, Scotland, and it was launched in June 1906, being able to make her maiden voyage to New York in September 1907. The company that owned the ship was the Cunard Line, which had been established in 1839 as the British and North American Royal Mail Steam Packet Company. Cunard had long been involved in transporting passengers, but also had a mail transport contract dated from 1840. This contract required that Cunard ships would be developed so they could easily be adapted to carry guns at time of war, and indeed 11 Cunard ships were requisitioned in the Crimean War in the 1850s. Some Cunard ships were also chartered by the British government in the Boer War in 1899–1901, adding significantly to the Cunard Line's profits during that period. However, in both these occasions there was very little risk to Cunard—the Russians not having a significant Crimean fleet

(the war being largely to stop this from happening), and the Boer Republics not having navies (and indeed both being landlocked).

Thus by the time of the construction of the *Lusitania*, there had been a long history of ships in general, and Cunard ships in particular, being requisitioned for the Royal Navy. There was nothing unusual in the British government providing low-interest loans to help Cunard build the *Lusitania*, nor in some of the crew also being members of the Royal Naval Reserve, and this was well known at the time of construction. Undoubtedly if there was a war, the *Lusitania* would serve as a troopship as many other passenger ships had done for the British at time of conflict. She could also be turned into an armed merchant cruiser, and indeed was officially listed as one, even though she was never fitted with these guns. The German government had a similar strategy to encourage their merchant navy to build liners and other large ships, although for them the nature of their war effort would be different from that of the British.

When war broke out in 1914, the British Empire depended on its control of the shipping lanes around the world. One hundred years earlier, this had led to the British defeat of Napoleon, and it was what had made the British Empire so prosperous. British domination of world trade and the ability of the ships of the Royal Navy and its allies to enforce their control of the seas were vital to the British war effort. It would ensure that food and raw materials would come to Britain, and also that Australian, New Zealand, Indian, South African, and Canadian soldiers, as well as those from other colonies, could easily get to the

---

Daniel Dow, Captain of the *Lusitania*

Prior to its fateful voyage, the captain of the RMS *Lusitania* was Daniel Dow. He had been born at Castle Bellingham, Ireland, on January 18, 1860, his father starting as a farm laborer and then becoming a carter. The family had then moved to Ireland, where his father had become a land agent.

Daniel Dow, the future ship captain, had started his career working for the ship *Baroda* in 1876. After five years with them, he married Margaret Cramb from Scotland, and they had a son. He also started working for Cunard. In September 1910 he was appointed as master of the *Mauretania*, and in March 1914 was appointed master of the sister ship, the *Lusitania*. Early in his command of it, he was involved in rescuing people from the *Mayflower*.

With the start of World War I, Daniel Dow, in charge of the *Lusitania*, was ordered by the British Admiralty to fly flags of neutral countries on the ship to try to prevent the ship from being attacked by submarines. Often it was the U.S. flag, largely because the ship was carrying many U.S. passengers. However, he found this extremely stressful and stopped sailing in the *Lusitania*. He later sailed the *Royal George* in the Indian Ocean, and went on to become involved in local politics in Liverpool. He died on January 20, 1931.

Western Front in France and Flanders. The Germans also recognized that they had to disrupt the shipping if they were to inflict damage on Britain. To this end, they used individual ships such as the SMS *Emden* in the Indian Ocean and also the submarines in the Atlantic Ocean.

Throughout World War I there were regular passenger liners traveling from the United States and Canada to Britain and back, and these brought civilians and people who were helping with the British war effort, including soldiers and nurses coming to Britain to enlist. The Germans recognized that these ships were clearly helping in the British war effort to the extent the Imperial German Embassy in Washington, D.C., took to advertising in newspapers that "travellers sailing in the war zone on ships of Great Britain or her allies do so at their own risk."

The idea that the British government used the civilian passengers on the *Lusitania* deliberately as "human shields" has often been raised. Certainly the *Lusitania* did carry weapons and ammunition, and even though this was down-played by the British at the time, it was known to military planners on both sides that all British ships crossing the Atlantic to the British Isles were likely to be bringing supplies. There were indeed gun mountings on the *Lusitania*, and the crew of the ship caught three Germans on board photographing these. They were arrested as spies and held in cells, and all three died when the liner was sunk (this accounts for the death toll being 1,201 rather than 1,198, as cited in many sources).

If it had been a conscious effort of the British to conceal the military supplies, the British government would have been using the passengers as a human shield. Instead they were so desperate for their war materiel to be carried across the Atlantic, the use of the *Lusitania* was done callously, because the ship was available and had large storage areas. The choice of the ship was therefore merely convenience rather than deliberate concealment or design.

Indeed the *Lusitania* was listed in *Jane's All the World Fighting Ships* as an auxiliary cruiser in its 1914 edition, along with the sister ship *Mauretania*. This clearly marked it as a British ship that could be used in the war effort. If the British government was deliberately using the passengers on the liner as human shields, the authorities would never have designated it as an auxiliary cruiser, the mere name indicating its potential use by the military.

At the time of the *Lusitania* sailing, the British war effort was desperately short of artillery shells. These were being used by the British artillery to shell enemy trenches, and were being fired at such a rate that production in Britain alone could not keep up with the demand. This situation was exacerbated with so many British factory workers serving in the armed forced. For this reason the British had sought to source more artillery shells and other explosives from the United States, and also from Canada. These shells would have been the ones transported on the *Lusitania*.

The greatest advantage of storing war materiel on the *Lusitania* was due to its design, which was also, incidentally, the reason why the ship would later to

sink so easily. The original liner was well designed by naval architects so that it had a number of water-tight compartments. Thus if one of them was punctured either by an accident, collision, or torpedo damage, the ship would not sink. The massive bulkheads protecting the compartment walls were so strong that they could cope with the pressures in times of crisis. However, the *Lusitania* was built before the tragedy of the sinking of the RMS *Titanic* in April 1912, so the lessons learned from the faulty design of the *Titanic* could not have been incorporated into the design of the *Lusitania*, although all ship crews were wary of icebergs.

Essentially to use more of the below-deck space for storage—in the case of the voyage in May 1915 of artillery shells and other war materials—some of the watertight compartments were merged by the removal of some of the bulkheads. Also to aid in the speed, important changes had been made in the design to allow the coal to flow more easily from one of the compartments to the engine. It was this adaption that had led to a much larger storage area becoming available. When the ship set sail, the compartments would be separated, but as the coal was needed, the connecting doors would be opened, allowing coal to flow from one area to another. In theory the doors could be closed, but with a large tonnage of coal forcing them open, it would not be possible to close them when at sea—and not without a large amount of work on land, if there was still considerable coal in storage.

The adaption meant that if a torpedo did hit, or if ship were damaged in any other way, flooding of that part of the ship would lead far more easily to the vessel sinking. This alteration clearly endangered the ship far more than had been the case, and Diana Preston (2002) argues in her book that it was one of the major causes for the *Lusitania* sinking when it finally was hit by a German torpedo. There were also indications that there may have been a secret explosive powder carried on board the ship, which was the real cause of the second explosion, and some have suggested that the coal dust may have became flammable. That the torpedo might have hit the artillery shells or other war supplies could easily account for the rapid sinking, but even if this was not the case, the removal of some of the bulkheads was clearly going to be a problem.

Much is made of the fact that the *Lusitania* was secretly carrying weapons, but on its final voyage this fact seems to have been quite widely known. The arrest of the three German spies is important—it is known that some German Americans had made a previous voyage on the ship and were so worried that they did make contact with the Imperial German Embassy in the United States, which then placed its infamous advertisements in the *New York Times* alongside that of Cunard's advertising.

Certainly weapons were listed on its original goods manifest, although the number of them and what exactly they were is in some doubt. Indeed two separate cargo manifests were lodged by Cunard with the U.S. Customs authorities.

One was lodged before the ship had actually departed and the other afterward. This has led to accusations that Cunard was deliberately seeking to mislead the New York harbor authorities and the U.S. Customs Service, and there is probably some truth in this. British companies during the war often would submit conflicting inventories and manifests in order to confuse German spies thought to be operating in the New York docks, and British intelligence was aware of German Americans operating in a number of important positions in the United States. The military cargo on board the ship was no real secret—and indeed was noted in the *New York Times* on May 8. There were even authors claiming that it was carrying large supplies of gold, possibly as much as $5 million in the precious metal. Yet none of the authors who have accepted this have ever answered the question as to why the gold would be carried from the United States to Britain, rather than the other way around, which seems more logical.

At that time, *Jane's All the World Fighting Ships* and similar books were used to identify ships, and the British were keen on disguising the *Lusitania*, which, as mentioned before, was listed as an auxiliary cruiser in the book. To help confuse the issue, the name *Lusitania* on the bow of the liner had been painted out, as was the case on many other vessels during war. This would not change identification by submarines, who did not rely on the names—and would never be close enough to read them anyway. But it might briefly confuse spies in the docks, or at any rate create a situation of some doubt for them, although some might argue this would not have caused much of a problem as the who passengers were sailing had seen the advertisements in newspapers for the sailing of the *Lusitania* with their safety disclaimers, and a telephone call to the Cunard office would have confirmed this.

It also appears that the funnels of the *Lusitania* had been painted a type of dusky gray, almost black—previously they had been Cunard's colors of red, with a black stripe across the top. Although some books cite this as a fact, there is some doubt. Certainly the logbook of the U-20's Kapitänleutnant Walther Schwieger, on the submarine that sank the *Lusitania*, stated that the funnels had been painted black. Some authors have commented that this was self-serving and that Schwieger could easily have altered his logbook—and this is certainly possible. *Lusitania* enthusiast Eric Sauder (2005) spent a long time trying to get confirmation for the information in Schwieger's logbook. To complicate matters further, film footage and photographs of the *Lusitania* on earlier voyages have often been portrayed as being taken on her last voyage, and these (even though they are in black and white) suggest the Cunard colors were used earlier in the war.

George Henderson, who was six when he saw the *Lusitania* in the distance from the Old Head of Kinsale, remembered them as being black. And work by archaeologists to locate the funnels has been unable to answer the question conclusively. However, it does appear that Walther Schwieger's logbook was probably accurate. Able Seaman Leslie Morton, an 18-year-old lookout, in his account

of the *Lusitania*, does mention that he was detailed to paint the lifeboats with gray paint, which again suggests that if this was done, the funnels might also have been painted. The most conclusive evidence emerged from a letter by Alice Loynd, a passenger on the *Lusitania*, who mentioned that in letters "we found that the *Lusitania* has been painted a different color. The Funnels are black, and the boat where it was white is now . . . grey" (Sauder 2005).

However, whether or not the funnels were painted, certainly to help confuse submarines, a compass platform had been added to the top of the bridge, and another was also added between the first and second funnels. There was also a pair of luggage cranes added to the aft deckhouse. Combined with these changes, and the fact that the *Lusitania* flew no flags—British or otherwise—during her last voyage provides evidence to back up the argument that the Germans knew that the British were using the *Lusitania* and other vessels to carry supplies, and the British therefore set out to conceal the *Lusitania*'s identity. Indeed on a previous recent trip across the Atlantic, German Americans traveling on it were concerned for their safety, and on a transatlantic voyage in early March 1915, a month after the Germans had declared all the seas around the British Isles to be a war zone, later stating that any Allied ships in the area could be sunk without warning, the British Admiralty had given the *Lusitania* very specific instructions on how to avoid submarines and even sent two naval vessels to help escort the *Lusitania*, although this did not in fact happen as the *Lusitania*'s captain was worried that the approaching ships could be German and successfully evaded them.

Therefore, if the British had actually wanted the *Lusitania* to be able to be spotted and identified correctly as a passenger liner from the Cunard Line, well known for sailing back and forth across the Atlantic, with large numbers of civilians, Americans and others known to be on board, with these passengers serving as human shields, it would be totally counterproductive to disguise the ship at all. The name would have been left on the bow, and everything possible would have been done to show that the vessel was the *Lusitania* as it could then be clearly identified by German U-boats as a civilian vessel and one they would recognize could potentially be carrying many U.S. civilians. The changes in the profile of the ship were to disguise its appearance sufficiently to confuse any German submarines, which the *Lusitania* could then easily outrun.

The British and the Cunard management clearly regarded the *Lusitania* as safe because, up to that point, no large passenger liner had been sunk. In spite of the disguise, it would be quite evident to all that the vessel was a passenger liner, even if submarines would have—it was hoped by the British—some trouble identifying her correctly. Safety essentially lay in the speed of the *Lusitania*, which was capable of going 21 knots (39 km) per hour. This was less than her previous prewar speed of 26 knots (48.2 km) per hour, as boiler room 4 in the vessel had been shut down to help reduce costs. However, it was almost twice the speed of submarines, which at that time could travel at only 11 knots (21 km) per hour.

Some authors have confused the issue slightly more. This is because there are a number of theories over the actual sinking of the *Lusitania*. Indeed it should be remembered that the vessel had almost completed her voyage across the Atlantic without incident when she was sunk off the coast of Ireland. Because of fog in the region, the *Lusitania* had actually slowed down to 18 knots. Certainly the weakening of the bulkheads to allow the coal to be used more easily was a major problem when the torpedo hit the ship. However, queries are raised over why the initial blast on the ship was so significant. Raimund Weisbach, the torpedo officer on the U-20 submarine that fired the torpedo, felt that the explosion was far larger than he had expected it would be. This has led to the theory that the torpedo hit some of the war materiel, including explosives—possibly secretly stored gunpowder—and that this ignited and then exploded, causing major damage. Apart from Weisbach's account, there are also the memories of many of the surviving passengers who have stated that there were two explosions on the ship, the first being cited as the one when the torpedo hit the *Lusitania*, and the second initially thought to have been a second torpedo. The log of the U-20 shows only one torpedo was fired, but there were again suggestions made that the log had been altered. Firing two torpedoes into a civilian liner would clearly cause it to sink quickly and result in far more deaths than would otherwise be the case. Therefore, there are arguments that the logbook is wrong. However, what many authors now suggest is that it was the explosives on board that ignited, causing the second, and more damaging, explosion, which led to the ship sinking so quickly. This seems plausible, although some authors cite slightly different reasons for this happening.

The sinking of the *Lusitania* with so many civilians and crew losing their lives did result in much criticism of Germany and what was seen as the German callous attitude to civilians, coming after the propaganda about their attacks on Belgian women and children, which had done so much to incite anger against Germany earlier in the war, resulting in many contemporary accounts referring to the Germans as the "Huns." The British government did not mention the carrying of artillery shells or cartridges, although Herman Winter, the assistant manager of the Cunard Line, was quoted in the *New York Times* on May 10, 1915, as stating that there were 4,200 cases of gun cartridges on the vessel, although he did state that these were for "small arms, packed in separate cases." Thus the storage of weapons of sorts was known soon after the sinking and published in a major world newspaper.

In summary, therefore, it seems that the *Lusitania* essentially only sunk after having slowed down toward the end of an otherwise uneventful passage. The British had used it for bringing war materiel from the United States, and the *Lusitania*, designated as an auxiliary cruiser, was clearly meant for this kind of work. However, the British had gone to some effort to disguise the appearance of the ship to make the German submarine crew think, if they did see it,

that the liner was not the *Lusitania*, but rather another unidentified vessel sailing across the Atlantic. It was the change in the bulkheads that caused the ship to sink so quickly with such a high loss of life. This action by the British government, and also by Cunard, was certainly reckless, and indeed callously put at risk the civilians, and indeed all on board, but it was not a deliberate attempt by the British to use civilians as human shields.

References and Further Reading

Bailey, Thomas Andrew, and Paul B. Ryan. *The Lusitania Disaster: An Episode in Modern Warfare and Diplomacy*. New York: Free Press, 1975.

Hoehling, A. A., and Mary Hoehling. *The Last Voyage of the Lusitania*. London: Longmans, Green, 1957.

Molony, Senan. *Lusitania: an Irish tragedy*. Cork: Mercer, 2004.

O'Sullivan, Patrick. *The Lusitania: Unravelling the Mysteries*. New York: Sheridan House, 2000.

Preston, Diana. *Wilful Murder: The Sinking of the Lusitania*. London: Doubleday, 2002.

Sauder, Eric. "The *Lusitania* Disguised? Unraveling the Mystery." 2005. http://marconigraph.com/lusitania/mfa_funnel.html (accessed June 18, 2010).

Simpson, Colin. *Lusitania*. London: Longman, 1972.

6

Woodrow Wilson's neutrality in World War I was so blatantly pro-British that he forced the Germans into attacking U.S. shipping.

PRO	Walter F. Bell
CON	Justin Corfield

PRO

The position that Woodrow Wilson's neutrality during the early years of World War I was largely responsible for the German attacks on U.S. shipping was widely popular among scholars and writers in the 1920s and 1930s who were influenced by the post–World War I disillusionment Americans felt toward Europe and which experienced a revival in the 1960s during the Vietnam War. It derived from the view that the U.S. position in the war on the Allied side was unnecessary and served no constructive purpose and that the administration of President Woodrow Wilson could have avoided war with Germany had the president and his advisers been as firm with Britain in asserting U.S. rights to trade with Germany as they were with the Germans regarding submarine warfare and the safety of Americans on the high seas. Wilson was inherently sympathetic to the cause of Britain and the other entente belligerents (the Triple Entente included Britain, France, and Russia). Because of his anglophile outlook, the president was vulnerable to (or supportive of) British efforts to lure the United States into the war. Because of his support for the Allies, Wilson lacked credibility with the Germans as a potential mediator of an end of the war—a perception that undermined the president's own efforts to bring about an end to the war before the United States was drawn in.

Wilson's pro-Allied (particularly pro-British) bias developed in the context of the growing Anglo-American friendship in the late 19th and early 20th centuries. These ties were cultural and economic as well as diplomatic. The United States and Britain had cooperated on a number of international issues. Britain had supported the United States in the Spanish American War, including American acquisition of the Philippine Islands; had allowed the United States to take the lead role in constructing the Panama Canal; and backed continuing U.S. interventions in the affairs of Cuba and other nations in Central America and the Caribbean. Advocates of overseas imperial expansion in the United States saw Britain as a kindred

127

spirit. At the same time, there were growing suspicions in these same circles of German expansionism.

There were also powerful ties linked to common language, cultures, and political systems. The U.S. political system derived from that of Great Britain as did the U.S. justice system. Intellectually, Americans identified with the writings of Shakespeare and Charles Dickens, the poetry of Byron and Shelley, and the political philosophies of John Locke, John Stuart Mill, and Jeremy Bentham.

These diplomatic and cultural ties were reinforced by close economic ties. Britain and the United States had been each other's best trading partners since the end of the War of 1812. British capital had financed a large portion of U.S. economic development in the 19th century. Moreover, the largest American banks—particularly J. P. Morgan—had close financial relationships with their counterparts in Britain and France.

President Wilson, a historian and political scientist before entering politics, had expressed admiration for Britain's constitutional system and expressed his beliefs that the United States and the United Kingdom shared common democratic values. When faced with the outbreak of war in August 1914, he referred back to his own historical writings to find a parallel with the War of 1812 with Britain—a war that Wilson believed had been forced on President James Madison by popular pressure and extremists in Congress and had been contrary to U.S. interests. He viewed Napoleon in 1812 and Wilhelm in Germany in 1914 as threats to civilization and to U.S. national security.

This analogy with the War of 1812 would govern Wilson's maritime rights policy throughout the period of U.S. neutrality from August 1914 to April 1917. It was based on three assumptions: (1) that U.S. security would be endangered by a German victory; (2) that the maintenance of U.S. maritime commercial rights did not justify a break with Britain; and (3) the Wilson administration must not allow public opinion to govern U.S. foreign policy.

Wilson's views were reinforced by his closest advisers. Edward M. House, although he did hold a position in the administration, was the president's closest friend and political confidant and shared Wilson's views concerning the need to conciliate the British. Most of the president's cabinet backed his decision, particularly Secretary of the Treasury William Gibbs McAdoo (also Wilson's son-in-law). The U.S. Ambassador in London, Walter Hines Page, was a close friend of the president and was stridently sympathetic to the British cause. Both House and Page had a cordial relationship with British Foreign Secretary Edward Grey, who would prove to be adroit in heading off stronger U.S. measures against the blockade. Another high administration official with pro-British sentiments was State Department Counselor Robert Lansing, who would formulate the legal framework for Wilson's neutral rights stance.

The only exception to the prevailing pro-British sentiment among the senior administration officials was Secretary of State William Jennings Bryan. Bryan

was a pacifist and staunch advocate of strict neutrality. Bryan was probably the most powerful political figure in the Democratic Party. Wilson had appointed him secretary of state as repayment for supporting Wilson for the democratic presidential nomination in 1914. Generally unschooled in international affairs, Bryan had become an object of ridicule in the administration, and his influence declined steadily after the onset of the war.

Early in the war, Wilson was determined to keep the United States out. In 1914 and early 1915, the greatest danger, in the view of the president and his advisers, lay in a confrontation with Britain over the rights of Americans to trade freely with both sides. The initial sticking point was the Declaration of London. A product of the London Naval Conference of 1908–1909, the powers intended it as a definitive set of rules to govern the maritime relations of neutrals and belligerents. It defined items that were considered contraband of war and therefore subject to seizure by a blockading belligerent. The declaration restricted the definition of contraband to items—such as arms, ammunition, and the raw materials required for their manufacture—that were clearly defined. The declaration had wide support in the United States but had been hotly debated in Britain, even though British negotiators at the conference had agreed to it. None of the signatories to the declaration had ratified it. When the war broke out, it was not in force.

Although the Wilson administration sought to gain the promise of both camps to support the Declaration of London, it soon became clear that the British, who controlled the seas, would not abide by it. A British Order-in-Council issue on August 20, 1914, proclaimed that food and oil (both items excluded by the Declaration of London) were "conditional contraband" and subject to seizure. Furthermore, the British Admiralty announced that the Royal Navy would mount a "distance blockade" that would close the North Sea approaches to the Netherlands and Germany and would implement the Doctrine of Continuous Voyage, which would permit them to take ships bound for neutral ports in countries neighboring Germany (mainly the Netherlands) if they believed the ultimate destination was Germany.

These measures clearly disregarded established international law and constituted a blatant violation of maritime neutral rights. Yet Wilson and his advisers, in the end, acquiesced to the British measures. The State Department prepared a protest note that bluntly cited Britain for violating international law. When they received the note, however, Wilson and House temporized and decided to pursue informal negotiations to try and modify the blockade. Wilson's main concern, again, was to avoid a major controversy with the British government.

Because of this concern to avoid controversy with Britain, the Wilson administration's foreign economic policies were also skewed toward the Allies. In August 1914, the U.S. economy was in a depression, which the outbreak of war threatened to make more acute. The threat of disruption to the cotton trade, which was vital to the economy of the U.S. South (a major part of the democratic

Wilson's political constituency), was particularly severe. As part of their August Order-in-Council, the British had restricted cotton shipments not only to Germany but also to Dutch, Norwegian, and Italian ports. Cotton prices tumbled and enraged cotton growers who demanded that Wilson take action against Britain. In response, the British War Cabinet, at the strong urging of Foreign Secretary Grey, agreed to purchase all of the surplus cotton at market price. As the war continued, Allied orders for cotton pushed exports above prewar levels. The threat of a confrontation over cotton had been defused.

The U.S. economy also benefited from Allied purchases of U.S. raw materials and munitions. Beginning in early 1915, the value of war-related exports to Britain and her allies experienced a major increase. The value of explosives shipped from the United States to Britain and France rose about fifteenfold ($2,793,530 to $32,197,274) between March and November 1915. Similarly, the value of exports of iron and steel more than quadrupled during the same period (Tansill 1938: 55). The ripple effect of these war orders in terms of jobs created and growth in business and banking sectors was decisive in moving the United States out of a depression and into an economic boom. These orders and the benefit to the U.S. economy were instrumental in binding the United States to the entente camp.

Benevolent neutrality toward Britain was also reflected in the Wilson administration's stance toward loans to Britain and France. At the war's onset, Secretary of State Bryan (who desperately wanted to keep the United States out of the war) urged Wilson to prohibit private loans to the belligerents. Bryan feared that a financial interest in the success of a debtor at war would generate pressure to enter the war to protect private financial interests. Initially, Wilson shared Bryan's concerns and imposed a ban in August 1914. As the war continued, however, pressure to reverse the ban grew as British and French credits dwindled

---

Senator Gerald Nye and the Neutrality Acts

Gerald Nye, one of the main proponents of the Neutrality Acts passed between 1935 and 1937, was a U.S. senator from North Dakota. Born in 1892 in Wisconsin, his father was the owner of a small newspaper and when he was six, Nye remembered seeing soldiers leaving for the Spanish American War. The family members were strong supporters of Robert La Follette and were against U.S. involvement in World War I. He reminded everyone that both his grandfathers had served in Wisconsin units during the American Civil War.

Initially Nye had contested a seat in the U.S. House of Representatives as a progressive republican in 1924. He failed in his bid, but in the following year when Senator Edwin Ladd died, Nye, having become editor of his newspaper, turned up to take down the governor's statement. He was shocked when the governor announced that he had nominated Nye to fill the vacancy. It was in Washington that Nye was appointed to the Special Committee on Investigation of the Munitions Industry.

and their ability to buy war materiel in the United States began to shrink. American bankers, led by the J. P. Morgan Company with the support of Undersecretary Lansing in the State Department, pressured the administration to lift the ban, citing the danger of the decline of exports and injury to the economy if the ban stayed in place. In August 1915, Wilson lifted the ban and stated that he would neither support nor oppose loans to belligerents—a step that in effect gave bankers the blank check they had sought. By not opposing private loans to Britain and France, Wilson had further sanctioned indirect support of the Allies.

President Woodrow Wilson with Secretary of State William Jennings Bryan, January 1913. Bryan, a pacifist and opponent of internationalism, resigned from Wilson's hawkish administration in 1915. (Library of Congress)

Pro-Allied sentiment in the United States was also bolstered by a highly successful British propaganda campaign. This operation cast the Germans as barbarians with no regard for international law who wantonly butchered civilians (including bayoneting babies) and ruthlessly exploited the resources of the areas they occupied. The centerpiece of this campaign was the report of the Bryce Commission on German Conduct in Belgium. Released in May 1915, the Bryce report detailed a long and highly sensationalized list of German atrocities, which included reports of the murders of thousands of civilians including women and children, the burning of hospitals, and thousands of rapes.

The Bryce report had far-reaching effects among the U.S. public and made a deep impression on President Wilson. The report was given credence by the reputation of its author—Viscount James Bryce—a respected British publicist who was known as a friendly critic of the U.S. political system. Bryce had many friends in the U.S. social and governing establishment, including Theodore Roosevelt. Wilson himself had a long-standing acquaintance with Bryce and had the highest regard for his opinions. Neither the president nor any other of his friends in the United States had any reason to suspect that either Bryce was now so senile that his ability to judge evidence was impaired or that his intense patriotism would make him a willing propaganda agent for the British government. Bryce's reputation as a scholar and his unimpeachable character seemed a safeguard against conclusions dictated by war hysteria.

The Bryce Commission's allegations were supported by reports from Ambassador Page in London and Minister to Belgium Brand Whitlock, both of whom were stridently pro-British and anti-German. Wilson and his advisers either ignored or dismissed reports received by the State Department from consular officers in Belgium and Germany that reports of German misconduct in Belgium were either exaggerated or false. The Germans themselves showed little sensitivity to public opinion in the United States. Incidents such as the execution of nurse Edith Cavell for espionage in 1915 unwittingly reinforced growing anti-German feeling among the U.S. public.

The greatest boost to anti-German sentiment in the United States, however, was the sinking of the British passenger liner *Lusitania* by a German submarine on May 7, 1915, with the loss of over 1,000 lives, including 128 Americans. The timing of the incident, coinciding with the release of the Bryce report, constituted a serious blow to German-American relations and a turning point for U.S. neutrality.

The torpedoing of the *Lusitania* took place against a backdrop of increasing German resentment over the United States' growing munitions trade with Britain and France and the failure of the Wilson administration to uphold established international law in the face of a British blockade, which had defined food as contraband and as a result caused increasing food shortages and hardship for the civilian population in Germany. Ambassador James Gerard in Berlin and the German ambassador in Washington, Count von Bernstorff, were warning the State Department that elements in the German government, led by the navy, were agitating for wider use of submarine warfare against Allied and neutral commerce on the high seas and for a submarine blockade around the British Isles. It would be difficult for the moderate leadership, led by Chancellor Theobald von Bethmann-Hollweg, to keep them in check unless the United States acted to force Britain to modify the blockade.

Despite these warnings and Bryan's efforts to reverse the president's course, Wilson and his advisers had not foreseen a confrontation with Germany over submarine warfare. Prior to the war, the submarine, as new technology, had not figured in the planning of either camp. Scholars of international law did not give the use of submarines against commerce much consideration. The law of nations had long mentioned the right of merchantmen to carry defensive armament. As submarines were vulnerable to even the smallest caliber of cannon and would be at risk if they were abiding by the rules of cruiser warfare, they were not considered useful as commerce raiders. The tactic of attacking a submerged submarine without warning had not been addressed. Early U-boat successes against British warships in the North Sea led many German naval officers, led by Secretary of State for Naval Affairs Admiral Alfred von Tirpitz, to recognize the submarine's potential as a threat to British commerce. Tirpitz and his followers pointed out that, as Germany had no other recourse with the Allies controlling the shipping

lanes, the navy should employ unrestricted submarine warfare against Allied shipping and neutral ships carrying contraband.

The first signs of trouble between Germany and the United States over the U-boat issue appeared in February 1915 when Germany declared a war zone around Britain. Any Allied ships within that zone were subject to being sunk by German submarines. It was inevitable that a British ship with Americans on board would be sunk, bringing on a confrontation. On March 28, the British steamer *Falaba* was torpedoed and sunk by a U-boat in the Irish Sea, killing 105 people including one American. The *Falaba*'s sinking shocked the president and his cabinet. A strongly worded protest note written by Robert Lansing condemned the sinking as a violation of international law and demanded that German submarines adhere to the rules of cruiser warfare. Despite heightened tensions, Secretary Bryan prevailed on Wilson to negotiate with the Germans. Even so, U.S. insistence on strict German adherence to the Wilson administration's interpretation of international law in contrast to the Wilson administration's comparatively mild reaction to Allied violations was a sure source of trouble in the future.

The sinking of the *Lusitania* on May 7, 1915, with the loss of 128 American lives, made clear to the president the dangers involved in his pro-Allied "benevolent neutrality." The Germans could not have picked a more high-profile target. It was one of the largest of Britain's passenger liners and a symbol of British power and pride. It sailed from New York on May 1 bound for Liverpool. Despite warnings from the German embassy in Washington that it would be subject to attack without warning and that neutral citizens traveled at their own risk, the liner had no escort and sailed with nearly 200 Americans on board. In addition, Bryan had warned Wilson that the *Lusitania* was carrying ammunition and begged the president to keep Americans off the ship. (There are allegations, never proven, that the British Admiralty, led by First Lord Winston Churchill, deliberately set up the *Lusitania* to be sunk in the hopes that the United States would enter the war on the Allied side.)

The Wilson administration's response to the crisis brought on by the *Lusitania*'s sinking illustrates the extent to which U.S. neutrality had slanted toward Britain. Despite Bryan's efforts to effect a compromise, the U.S. notes to the German government sounded a stern demand that Germany disavow the sinking, punish the officers responsible, take responsibility for the sinking, and pay an indemnity for the loss of American lives and property.

Wilson resisted Bryan's efforts to affect a compromise by prohibiting Americans' travel on belligerent vessels or to impose a ban on shipping war materials on any passenger vessels in exchange for a German agreement to refrain from attacking these ships. The president brushed aside German arguments that the fact that the *Lusitania* was carrying munitions and troops made it a legitimate target. Nor would Wilson link his demands for German concessions to any quid pro quo on his part to step up pressure on Britain to relax the

blockade (although he hinted informally to von Bernstorff that German agreement to give up the submarine would be an encouragement for him to do so).

The hard line Wilson took toward the Germans, including the threat to break diplomatic relations if the Germans did not meet his demands, caused Secretary of State Bryan to resign in protest in June 1915. Bryan's resignation reflected his frustration over his inability to move the president toward a more balanced neutrality. During the crisis, Wilson relied increasingly on the advice of the pro-Allied counselor Lansing, who drafted the U.S. notes. Wilson generally accepted Lansing's narrow definition of the law of the sea. Lansing's position that German submarines must follow the rules of cruiser warfare made no allowance for the nature of submarine warfare and put administration policy in a straitjacket.

Another crisis followed in August 1915 when a U-boat torpedoed and sank the British passenger steamer *Arabic*. Again the American reaction was to demand that Germany disavow submarine warfare. The crises surrounding both the *Lusitania* and *Arabic* incidents ended when Ambassador Bernstorff announced that German submarines would not sink passenger liners without warning and without providing for the safety of noncombatant passengers as long as the ships did not resist or try to escape.

Bernstorff's *Arabic* pledge, which he made without instructions from his government, effectively ended the first German-American crisis over submarine warfare. The German decision to back down reflected the temporary ascendancy of Bethmann-Hollweg and the moderates in the German government over the hardliners concentrated in the German navy. War with the United States had been averted at this time because Germany had not yet produced a sufficient number of submarines to make a blockade of Britain effective and because the Army High Command, which held the balance of power in Germany, did not want to risk war with the United States. The continuance of this uneasy balance of forces in Germany depended on the course of the war and the U.S. ability to induce Britain to relax the blockade.

Between August 1915 and the end of 1916, however, the position of the German moderates regarding unrestricted submarine warfare continued to deteriorate. This decline was due, in part, to rising popular discontent due to growing food shortages related to the blockade. High casualties and frustration with the stalemate on both Eastern and Western Fronts were triggering calls from the army for steps to bring about a quick military solution. Increasingly, the military and their right-wing civilian supporters saw the submarine as a panacea to their problems. They called for an unrestricted submarine war against Allied and neutral commerce that would starve Britain and France into making peace.

The only way the Wilson administration could undercut this trend in Germany was to take steps to induce Britain to relax the blockade and allow noncontraband goods and food into Germany. There were thorny issues that threatened the Wilson

administration's "benevolent neutrality." Americans were deeply upset over the seizure of U.S. merchantmen carrying cargo that the State and Commerce Departments regarded as noncontraband items. Particularly contentious was the British publication in July 1916 of a "blacklist" of companies that did business with Germany; their products were to be subjected to boycott by the Allies, and their assets in the Allied belligerent countries would be confiscated. Despite bluntly worded protests from the State Department, the administration never considered either embargoing the flow of war supplies to Britain and France or cutting off loans to Britain and France. The economic boom in the United States was too closely tied to trade with those countries, and public opinion, though opposed to U.S. entry into the war, was still predominately sympathetic to the Allies.

Wilson understood that a prolonged war would increase the possibility that the United States would be pulled in. The only way the president could see out of his dilemma was a compromised peace, with the United States as the arbitrator. Even on the issue of a mediated settlement, Wilson and his main emissary and confidant Colonel House viewed Britain as the main avenue for convening a peace conference.

The administration's main effort to induce Britain to support a U.S. peace initiative began in the winter of 1915–1916 when Wilson sent House to Europe to probe the belligerent's receptivity to possible peace mediation. House visited Berlin and London. The most significant development came in his talks with Foreign Secretary Grey, which resulted in the House-Grey memorandum. Drafted by Grey, the memo stated that the president was ready, on hearing from France and England that the time was right for a peace conference, to issue a public call for such a meeting. If the Allies accepted Wilson's offer and Germany did not, then the United States would "probably" enter the war against Germany.

The gist of this document was that the Wilson administration tied its hopes for mediating a compromise settlement of the war to the willingness of the Allies to follow the U.S. lead. Grey encountered intense opposition to the memo from both members of the British War Cabinet and from the French, and no peace message ever came from the Allies. Nevertheless, to many postwar critics of Wilson's neutrality policies, the House-Grey memorandum seemed designed to provide a path to war rather than a path to peace and a way out of Wilson's predicament.

Throughout 1916, Wilson continued to seek ways of moving the warring camps toward a peace conference and a compromise settlement. The centerpiece of these continuing efforts was his speech before the League to Enforce Peace on May 26 in which he stated the principles that should form a broad outline for a settlement and appealed to the Allies to call for a peace conference. The speech was basically a call to the British and France to act upon the understanding between Grey and House. Wilson also called for an association of nations to enforce a settlement.

This address reflected the president's continuing belief that acting through the Allies constituted the best means for U.S. mediation of a settlement. Wilson was careful not to say anything that might offend Britain or France. There was nothing concerning war aims or freedom of the seas. The Allies' failure to respond to Wilson's appeal led him to conclude that they were no more interested in peace than the Germans, and that independent efforts at mediation might be the best avenue to engineer a settlement.

By this time, it was too late. Neither side was interested in a compromise peace. Mounting casualties on both sides and the immense sacrifice of lives and treasure made a compromise peace increasingly difficult. Britain and France believed that their bargaining position was weak and that a peace conference would undermine their interests. In Germany, the faction favoring resuming submarine warfare and achieving military victory was gaining the upper hand. Led by Admiral Henning von Holtzendorff, the German hardliners argued that victory in Europe could be achieved even if the United States entered the war. Germany had to act immediately before the growing threat of starvation related to the blockade forced Germany to her knees. There was widespread feeling that with more U-boats now available and with Britain's domestic economy deteriorating, the time was now right to undertake an unrestricted U-boat campaign. They did not trust Wilson's motives in seeking mediation or his ability or willingness to end the blockade.

For the rest of the year, Wilson was preoccupied with the 1916 presidential election campaign. He was in a close campaign against Republican Charles Evans Hughes and did not want to undertake any measure that might adversely affect his chances for reelection. He did not fully realize that the political balance in Germany was shifting and that the United States was close to war with Germany.

It was in the context of the hardening of positions on both sides that Wilson's attempt at independent mediation in December 1916 failed. The initiative began in conjunction with a November 26 message from Bethmann-Hollweg transmitted through several neutral governments proposing that the Allies agree to send representatives to a conference to discuss peace terms. The German chancellor's action was probably an act of despair in the face of the political reality that the supporters of a military solution were gaining the upper hand in the government and that Germany was about to commit political suicide with the U-boat weapon.

Although reluctant to appear that he was cooperating with Germany (he had actually been drafting a message since three weeks prior to the appearance of Bethmann-Hollweg's message), Wilson felt that he had to act. On December 16, he issued an appeal for all belligerents to state their war aims. The Allies reacted negatively. The initial German reaction was conciliatory but right wing, and militarist elements in Germany almost unanimously did not want to use the president as a mediator. Wilson moved ahead with his celebrated "Peace without Victory"

address to the U.S. Senate on January 22, 1917, in which he offered conditions—mainly freedom of the seas—that he believed would attract the Germans.

It was already too late. The Germans did not respond to Wilson's proposals because they had already set their course. On January 8, in a meeting at the German Imperial Headquarters at Pless, Paul von Hindenburg and Erich Ludendorff informed the kaiser that the army now supported unrestricted submarine warfare. The conferees decided that the new campaign would commence on February 1, 1917.

With that decision made, the slide toward war between Germany and the United States became inexorable. With the sinking of a number of U.S. vessels in February, Wilson made the decision to break diplomatic relations with Germany at the end of that month. On April 2, Wilson asked Congress for a Declaration of War on Germany—an action Congress took on April 6, 1917.

In retrospect, a strong case can be made for the argument that Wilson's pro-British "benevolent neutrality" was doomed to fail and that the president, by believing that his pro-British orientation would avoid war, was guilty of monumental self-deception. Because he failed to consider alternative courses seriously, a confrontation with Germany was virtually unavoidable. From the beginning, Wilson relied on advice from senior members of his administration, as well as his close confidant Colonel House, who were pro-Allies and who, in many cases, had ties to economic interests with a stake in an Allied victory. Exports of munitions and raw materials to Allied nations and loans to Britain and France rendered U.S. prosperity dependent on the Allies.

Wilson and his advisers acceded to flagrant British violations of international law and the law of the sea. The British use of a legally questionable definition of food as contraband and the equally questionable doctrine of continuous voyage denied Germany imports of food and led to threats of a famine that forced Germany into a war-winning strategy with unrestricted U-boat warfare that was almost unachievable. In their application of international law, the president and his advisers used a double standard where they demanded that German submarines adhere strictly to rules of cruiser warfare while they looked the other way at British violations. Invoking international law against the U-boats was questionable in any case as the use of submarines as commerce raiders had never been addressed in the law of the sea.

With his anglophile orientation, Wilson also undermined his credibility with the Germans as an impartial mediator, thus crippling his attempts to bring about a compromise solution to the conflict that could have prevented a German-American confrontation. Ironically, Wilson wanted to keep the United States out of the war and believed that a peace dictated to Germany by the Allies would have disastrous long-term consequences, but he refused to take steps toward strict neutrality that would put the United States in a stronger position as a neutral and impartial mediator. It is a tragedy that the healing peace that Wilson

believed essential to civilization's future eluded him because of his administration's lack of neutrality.

References and Further Reading

Beard, Charles A. *The Devil Theory of War*. New York: Vanguard Press, 1936.

Coogan, John W. *The End of Neutrality: The United States, Britain, and Maritime Rights, 1899–1915*. Ithaca, NY: Cornell University Press, 1981.

Fleming, Thomas. *The Illusion of Victory: America in World War I*. New York: Basic Books, 2003.

Grattan, C. Hartley. *Why We Fought*. New York: Vanguard, 1929.

May, Ernest R. *The World War and American Isolation*, *1914–1917*. Cambridge, MA: Harvard University Press, 1959.

Millis, Walter. *Road to War: America, 1914–1917*. Boston: Houghton Mifflin, 1935.

Steffen, Dirk. "The Holtzendorff Memorandum of 22 December, 1916 and Germany's Declaration of Unrestricted U-Boat Warfare." *Journal of Military History* 68, 1 (January 2004): 215–24.

Tansill, Charles Callan. *America Goes to War*. New York: Little, Brown, 1938.

Tucker, Robert W. *Woodrow Wilson and the Great War: Reconsidering America's Neutrality, 1914–1917*. Charlottesville: University Press of Virginia, 2007.

CON

When World War I broke out, the U.S. government was keen to maintain its neutrality, and indeed it did not enter the war until April 6, 1917, when it declared war on Germany. However, there were some doubts, with claims that the government of Woodrow Wilson supported the British so openly that the Germans felt they needed to start attacking Atlantic shipping. From 1934 until 1936, the Special Committee on Investigation of the Munitions Industry, better known as the Nye Committee, met in the U.S. Senate to study the involvement of the U.S. government in the war, with many U.S. politicians arguing that Wilson was blatantly pro-British in his policies and actions so that German attacks on shipping became inevitable and thus precipitated the U.S. declaration of war on Germany.

The assassination of Archduke Franz Ferdinand in Sarajevo on June 28, 1914, was a major news story around the world, including in the United States. When war broke out four weeks later, with the Austro-Hungarian Empire declaring war on Serbia on July 28, this precipitated a series of alliances with

the Russians mobilizing their army and then Germany declaring war on Russia, then France, and invading Belgium; and Britain, declaring war on Germany; after which Japan entered the war, declaring war on Germany.

For most Americans in July and August 1914, the war held no great fascination. The wealthier Americans on the east coast (especially in New England) tended to support Britain and France through ties of ancestry and through France's help in the War of Independence, which had led to a close bond between the two countries, as seen in gesture of giving the Statue of Liberty to the United States. But this feeling was never universal among those on the East Coast of the United States, where there were many Irish Catholic families who detested the British and would support anyone who opposed them because of their rule over the whole of Ireland. There were also many German Americans who supported Germany, but also some who had fled Germany to escape its militaristic government or its domination by the Prussian state. Other German Americans wanted nothing to do with the fighting in Europe, the wars there being essentially the reason for their migration to North America.

Beyond the East Coast, many Americans in the Midwest and also on the West Coast had absolutely no interest in the events in Europe. Some were from Scandinavian countries, which remained assiduously neutral throughout World War I, and others whatever their ancestry, had moved to the United States because it represented the opposite, as they saw it, to the bickering countries of Europe arguing over borders, minority groups, enclaves, and extraterritorial jurisdictions.

On August 4, 1914, Woodrow Wilson issued a formal proclamation that the United States would remain neutral in the war. He noted:

> [T]he people of the United States are drawn from many nations, and chiefly from the nations now at war. . . . Some will wish one nation, others another, to succeed in the momentous struggle. It will be easy to excite passion and difficult to ally it. . . . Such divisions among us would be fatal to our peace of mind and might seriously stand in the way of the proper performance of duty as the great nation at peace, the one people holding itself ready to be a part of impartial mediation and speak the counsels of peace and accommodation, not as a partisan, but as a friend. . . . The United States must be neutral in fact as well as in name during these days that are to try men's souls. We must be impartial in thought as well as in action, must put a curb upon our sentiments as well as upon every transaction that might be construed as a preference of one party to the struggle before another.

He warned that if U.S. citizens did not scrupulously avoid taking sides in the war, Americans would "be divided in camps of hostile pinion, hot against each

other, involved in the war itself in impulse and opinion if not in action" (quoted from Maxwell 1914: 141–42).

Over the next two-and-a-half years, Woodrow Wilson certainly tried to keep the United States absolutely neutral in the conflict, although many Americans took sides, either emotionally or through actions to aid one side or the other. Indeed Herbert Hoover, himself of Swiss ancestors who had migrated from their home in Germany to North America in 1738, had been sent to Europe where he was involved in providing relief aid through the American Relief Committee in London. His Committee for the Relief of Belgium and Northern France helped provide aid for about 10 million people living under German rule and would gain him the reputation that would lead to his appointment as secretary of commerce in 1921 and election to the presidency in 1928.

And during those two-and-a-half years, there were many incidents and pressures that pushed Wilson toward entering the war on the British side, but he chose not to do so. In fact, his view on involvement in the war naturally changed as the nature of the war itself changed. Within those 32 months, Wilson's own views changed. By May 1915 and the sinking of the *Lusitania*, Wilson was increasingly pro-British, but he soon became angered by the British government and its general attitudes that by the middle of 1916, Wilson was much more in favor of continued neutrality. These shifts in policy would certainly not have been the case if he were doing everything he could to bring the United States into the war on the British side.

From an overview, there were strategists who argued that a British or French victory would be far better for the United States. Both these countries were considered "content." They had large empires and were basically "satisfied" with the situation in the world at large. This allowed the United States, through the Monroe Doctrine and other policies, to dominate the Americas because Britain and France had no interest in upsetting the situation. By contrast, a victory for Germany in the long term would result in victory for a "discontented" nation, which might then strive to have colonies. These are certainly viewpoints that were later raised by R. E. Osgood in his book *Ideals and Self-Interest in American Foreign Policy* (1953). This attitude might be cynical, but it also belies the fact that the Americans were also keen for their merchant navy to overtake that of Britain, as looked possible until the American Civil War and the actions of the Confederate raiders had changed the situation in the 1860s. Thus a war between Britain and France, against Germany, leading to all three countries exhausting themselves, might prove to be in the United States' long-term interest.

The initial German invasion of Belgium and France, and the atrocity stories run by British newspapers, certainly angered many Americans, as they were later shocked by the execution of the British nurse Edith Cavell, who was shot as a spy by the Germans in October 1915. The question in the minds of many Americans was with whom did they identify during that period of the war. There were still people in the United States who remembered the destruction of

the American Civil War, and many more who heard about it from their parents and grandparents. Yet for the United States to enter the war on the British and French side meant that it would be allying itself with Belgium and Russia, a situation that some Americans would have found untenable.

Belgium had a terrible record of human rights abuses in the Congo (modern-day Democratic Republic of Congo), which had been unmasked by Edmund Dene Morel in 1900 and brought to the attention of even more people around the world by Joseph Conrad in *The Heart of Darkness* (1902). Up until 1908 the king of Belgium directly ruled Congo as the Congo Free State, but international pressure forced a change with the creation of a colony controlled by the Belgian Parliament. The German invasion of Belgium, until the horror stories of the atrocities by German soldiers, did little to win sympathy for the Belgians. And similarly for Russia, many Polish and Jewish Americans would not support a war that would help maintain the Russian Empire from which so many of them had fled after generations of persecution. The United States in 1914, as it had done for so many years before, and as it does today, prides itself on being a focus for democracy in the world, and Wilson would certainly not have relished entering a conflict to support dictatorships, which, if they had entered World War I in 1914, they would inevitably have done.

Many pointed to the increased trade between Britain and the United States, especially in armaments, as being an important factor that helped sway Wilson. Father Charles Coughlin later drew parallels between the arguments for war and the profits that would be made by the emerging military-industrial complex. Certainly the U.S. economy did do well during the first years of the war, selling vast amounts of armaments and other products to the British. They also extended lines of credit, and as an example, exports of explosives that were worth $6 million in 1914, rose to $467 million two years later. However, the simple argument that Wilson was anxious to drag the United States into the war to help trade balances is wrong because, while it could do well *selling* weapons to the British, once it entered the war itself, the U.S. government would incur the costs of furnishing and equipping an expeditionary force at great expense to the U.S. taxpayers. The companies associated with the military-industrial complex may have wanted war, as Coughlin argued, but that did not mean that Wilson did. And during the late 1920s and early 1930s, these companies and interest groups, rather than Wilson, became the focus of the criticisms that would lead to the introduction of the Neutrality Acts of 1935, 1936, 1937, and 1939.

There was more trade with Britain than with Germany during the first 32 months of World War I. This was not engineered by Wilson either because he supported Britain or because he wanted to encourage attacks on shipping so that war might start. It was dictated by history—there had always been more trade between the United States and Britain than between the United States and Germany, mainly because of geographic circumstances. The British had imposed a naval blockade on Germany, which, because of its position, was unable to use the North Sea ports or

those captured from Belgium. Most of their navy was unable to "break out" from the encirclement by the Royal Navy, as seen by the Battle of Jutland and other engagements when they tried so to do. Thus for any U.S. companies to trade with Germany, they would have to get through the British naval blockade, and this in itself would cause tensions. The use of mines by both sides complicated things further and made it even more hazardous for any ships to try to reach German ports.

When the German government announced, on February 4, 1915, that they had declared the waters around the British Isles to be a "war zone," it meant that they could sink any enemy merchant vessels at sight. Six days later, the U.S. government stated that the Germans would be held to a "strict accountability" for any U.S. lives or property lost. Thus all Americans had, as they had since the start of hostilities, been advised to use ships from neutral countries. The situation was complicated because the British were not averse to "reflagging" some of their ships with neutral flags. Again this was not of Wilson's doing, but it did lead to increased tensions with Germany.

In previous conflicts, enemy ships regularly attacked merchant ships of their rivals, as was common with the Confederate raiders in the American Civil War. However, the Confederates and others beforehand generally took the crew of the ship they were sinking as prisoners, who were then released in a neutral port. For a submarine, it was impossible to take the crew of a ship they were sinking, and even more so if this was a passenger liner. Furthermore, the submarines were nowhere nearly as efficient as the U-boats of World War II, and the British instructed their ships to fire on or ram any German submarine. With many British merchantmen armed, the Germans claimed that changes in the rules of international law meant they could sink foreign vessels on sight, and without warning. This happened with the *Lusitania*, which went down with the loss of 1,201 men, women, and children. In response to this, on June 10–11, 1915, Secretary of State William Jennings Bryan did ask in statements published in the *New York Times* why U.S. citizens continued to travel on British ships. As he wrote, "The question is not whether an American citizen has a right, under international law, to travel on a belligerent ship; the question is whether he or not, out of consideration for his country, if not for his own safety, avoid danger when avoidance is possible" (quoted from "Mr. Bryan's Defense" 1915). It was a clear indication that the Wilson administration, while obviously horrified by the sinking of the *Lusitania*, was still keen on avoiding war.

In response to protests from Wilson, the Germans said that they would try to prevent anything similar from happening again. However, in March 1916 the channel steamer *Sussex* was torpedoed in the English Channel. In spite of its name, it was a French ship, and it managed to reach port with none of the U.S. passengers killed, although some were wounded. Robert Lansing who was then secretary of state, urged the United States to use this as an opportunity to declare war on Germany, but Wilson rejected this.

Woodrow Wilson campaign van, 1916 presidential election. Among the slogans, "Who keeps us out of war?" (Bettmann/Corbis)

In fact rather than deciding to support the British in the war, Wilson had been keen to try to prevent war. Throughout late 1916, Wilson was involved in peace moves, which might have succeeded because the Allies were suffering badly from the war. The French had failed to break the Germans at Verdun, and in the Battle of the Somme, the British had, likewise, not been able to break the German lines, suffering massive casualties. On the Eastern Front, the Russian army was in turmoil and being driven back on many fronts, with the Germans in control of Poland and a considerable part of the Baltic, resulting in massive tension in the Russian capital of Petrograd. So there was certainly a possibility of the various sides at least considering a ceasefire and a peace conference.

One of the influencing factors on Wilson's policy during this time was undoubtedly the presidential election in November 1916. Theodore Roosevelt had decided not to stand and split the republican vote, and this left Wilson against Charles Evans Hughes in what was clearly going to be a close election result. Wilson's democrats used the slogan "He Kept Us Out of War," while Hughes campaigned on the slogan of "America First, America Efficient." The end result was Wilson getting 9,126,868 votes to 8,548,728 for Hughes, with Wilson securing 277 Electoral College seats to 254 for Hughes. Wilson won California by less than 4,000 votes, and a loss there would have cost him the victory. Although Wilson won and had a majority of 53 to 42 in the senate, he only had a slim majority in the House of Representatives (216 to 210). As a result Wilson remained cautious.

The Black Tom Explosion and the End of Isolationism

One of the causes of U.S. involvement in World War I was the Black Tom explosion on July 30, 1916, in the harbor of Jersey City, New Jersey. It was a major munitions dump, and in spite of the United States being neutral, it was allowed to sell war materiel to the Allies.

The explosion took place just after 2 A.M., and it set off a series of fires, causing property damage valued at the time at $20 million. The explosion sent shrapnel in various directions and even damaged the Statue of Liberty. It is believed that seven people were killed, with hundreds suffering small injuries.

The bombing was connected to a Slovak migrant named Michael Kristoff who had the help of two guards of German ancestry. The event, and especially the damage to the Statue of Liberty, estimated at $100,000 and led to access to the torch being closed to the public—led to increased anti-German feeling in the United States.

Although Wilson had hoped that his moves for peace might help the Allies, in the end, he was upset that the British were reluctant to take up his idea. His dislike of the British position was increased when, in July 1916, the British had issued their "blacklist" of companies that had been involved in trade with Germany or had German connections. Wilson saw this as a deliberate attempt to interfere in U.S. domestic economic policies, and he wrote to Colonel Edward House, "I am . . . about at the end of my patience with Great Britain and the Allies. This blacklist is the last straw" (Jones 2001: 79). The result of this was that Wilson decided to go back to his original declaration of August 1914 in which the U.S. government would reflect total impartiality in the war. In December 1916, the Germans stated that they were interested in a negotiated settlement, but the new British Prime Minister David Lloyd George was against this. He felt that only a defeat of Germany could lead to peace. A ceasefire with a strong Germany would lead to another arms race, and then another conflict.

While this was happening, the U.S. government was also bolstering its military potential so that if it did become involved in World War I, it would do so from a position of strength, and likewise if the war ended, the United States would be able to defend itself. Thus the U.S. navy started construction of a number of battleships. This in itself was significant because if the U.S. government were preparing to enter the war in support of the British, then smaller and faster naval vessels that could chase German submarines would see the U.S. taxpayers' money spent more effectively. By contrast the battleships could challenge the British fleet.

In February 1917, the German ambassador to the United States left, and the U.S. ambassador to Germany, James W. Gerard, was recalled (or technically asked to leave by the Germans), severing diplomatic relations. But Wilson was

still keen on avoiding war. Indeed Edward House noted in his diary: "The president was insistent that he would not allow it [German actions] to lead to war if it could possibly be avoided. He reiterated his belief that it would be a crime for this Government to involve itself in the war to such an extent as to make it impossible to save Europe afterward" (quoted from Hoover 1958: 4). Certainly Wilson was more anti-German at this point, and House added that Wilson "spoke of Germany 'as a madman that should be curbed.' I [House] asked if he thought it fair to the Allies to ask them to do the curbing without doing our share. He noticeably winced at this, but still held to his determination not to become involved if it were humanly possible to do otherwise" (quoted from Hoover 1958: 4). What changed Wilson was that Germany continued to attack merchant ships. On February 25, 1917, the *Laconia*, another Cunard liner, was sunk, with the loss of two U.S. women. It came only a day after the British had given the U.S. government the text of the Zimmermann telegram by the new German foreign minister. Arthur Zimmermann had taken over as foreign minister when the German High Command insisted on unrestricted submarine warfare as the only possible way of winning the war, after Gottlieb von Jagow had resigned. In the telegram, intercepted by the British, Zimmermann offered the Mexicans Texas, Arizona, and New Mexico if they entered the war. The suggestions were specious, but the publication of the telegram did increase the level of mistrust that Wilson felt for the Germans.

It was the Russian Revolution of February–March 1917 that would lead to the U.S. involvement in the war in two ways. First, the collapse of the Russian army and their front meant that the Allies were in danger of losing the war because Germany could send more of its soldiers against the British and the French in France and Flanders. This made it, for the British and French, all the more important for the United States to enter the war. At the same time it also meant that if the United States entered the war, it would now be doing so to support democracies (Britain, France, and now Russia) against dictatorships.

However, it was the continued attacks on the shipping lanes by German submarines that would to lead to the U.S. declaration of war. On April 6, 1917, the House of Representatives by 373 to 50, and the Senate by 82 to 6, voted "That the state of war between the United States and the Imperial German Government which has been thrust upon the United States is hereby formally declared." Wilson then signed this into law on the same day. He stated that the war was not an ordinary war, but a crusade "to make the world safe for democracy."

Thus it can be seen that Wilson was keen on avoiding war, and although he certainly did sympathize more with the British than with the Germans, both his public statements and also his actions, as well as his private comments, show that he did try to steer the United States away from intervention in what was then known as the Great European War, and subsequently as World War I.

References and Further Reading

Albion, R. G., and J. B. Pope. *Sea Lanes in Wartime*. New York: Norton, 1942.

Buehrig, E. H. *Woodrow Wilson and the Balance of Power*. Bloomington: Indiana University Press, 1955.

Devlin, Patrick. *Too Proud to Fight: Woodrow Wilson's Neutrality*. London: Oxford University Press, 1974.

Hoover, Herbert. *The Ordeal of Woodrow Wilson*. Washington, DC: Woodrow Wilson Center Press, 1954.

Hunt, Michael H. *Crises in U.S. Foreign Policy*. New Haven, CT: Yale University Press, 1996.

Jones, Howard. *Crucible of Power: A History of U.S. Foreign Relations since 1897*. Lanham, MD: Scholarly Resources, 2001.

Lovell, S. D. *The Presidential Election of 1916*. Carbondale: Southern Illinois University Press, 1980.

Maxwell, Charles. *Causes and Battles of the Colossal War in Europe*. NP: W. J. Hainsfurther, 1914.

"Mr. Bryan's Defense." *Current History* (July 1915): 645–47.

Mulder, John M. *Woodrow Wilson: The Years of Preparation*. Princeton, NJ: Princeton University Press, 1978.

Osgood, R. E. *Ideals and Self-Interest in American Foreign Policy*. Chicago: University of Chicago Press, 1953.

Powaski, Ronald E. *Toward an Entangling Alliance: American Isolationism, Internationalism, and Europe, 1901–1950*. Westport, CT: Greenwood, 1991.

Pratt, Julius. *A History of United States Foreign Policy*. Englewood Cliffs, NJ: Prentice-Hall, 1965.

Seymour, Charles. *American Neutrality 1914–1917*. New Haven, CT: Yale University Press, 1935.

Tansill, C. C. *America Goes to War*. Boston: Little, Brown, 1938.

Tuchman, Barbara W. *The Zimmermann Telegram*. New York: Viking, 1958.

7

Mahatma Gandhi would not have been a world leader without the influence of Rabindranath Tagore.

PRO   Rajini Pani
CON   Rajshekhar

PRO

As contemporaries and leading torch-bearers of India's freedom struggle in the 20th century, Rabindranath Tagore (1861–1941) and Mahatma Gandhi (1869–1948) generate not merely academic interest among commentators, but also controversy and debate on a range of issues. Gandhi, as the world remembers him, was an apostle of nonviolence, the architect of India's freedom, and even revered in some quarters as an incarnation of the Buddha. Tagore, in comparison, registers several million hits less on any modern-day Internet search engine, in spite of being a poet laureate. Although Gandhi stands among the world's illustrious leaders, Tagore is romanticized as a singer of songs and a dreamer of dreams from Bengal. He lives more in Indian memory and less in that of the world's, as India's and Asia's first Nobel laureate. But supported by relevant historical documents, an objective analyst is bound to catapult Tagore from unwarranted relegation to the position of a worthy forerunner to Gandhi and Gandhian policies.

Tagore was born before Gandhi and was, like him, a citizen with a conscience, a patriot, a campaigner for social justice, particularly for women and children, and one who had infinite faith in the good of mankind. In short, Tagore was, as critics have acknowledged, far more than a mere poet, just as Gandhi was more than a mere politician. Hence it would be appropriate to begin with the premise that although the coming of Gandhi was said to be like a thunderbolt out of the blue, Tagore's influence before him was like the coming of dawn in the mountains that permeated the thinking of a generation that produced philosophers and fighters for India's freedom. It is therefore more important to understand Tagore's side of the now famous Gandhi–Tagore debate.

Older to Gandhi by eight years, Tagore was, sociologically speaking, a product of history, positioned by the times to initiate pioneering projects and actions, from which Gandhi learned. The sheer volume and diversity of Tagore's creative life, spanning 60 of his 80 years, created a legacy, which Gandhi and an independent India proudly displayed. Admirers of Gandhi would therefore find it difficult to deny equal star status for Tagore as a poet, philosopher, artist, and educator.

147

With 28 volumes of literary works, 2,500 songs, and 2,000 drawings, this legend inspired revolutionaries in Bengal, a state considered to be the eyes and ears of the Indian freedom movement.

In order to establish that Gandhi would not have been a world figure without the influence of Tagore, it would be necessary to illuminate those beams of over-arching influence that led Gandhi to Tagore. Both of these men became important for India at one of the most crucial stages of the fight against colonialism. The French philosopher Romain Rolland (1924) aptly sums up their position as two great river-like souls overflowing with the divine spirit. Tagore's influence on Gandhi was subtle of expression, gentle, yet powerful. This was by no means easy, as Gandhi's style and response had a rustic resistance about it, almost as rough as the homespun clothes he wore. Tagore was spontaneous in his state-ments, foreseeing the unpredictable and always ready to accommodate the unfore-seen. Gandhi's views, in contrast, were prescriptive to the point of being didactic. But what mattered more was Tagore's stance of carrying an argument to its logi-cal end, with conviction. Gandhi was often driven to learn from Tagore. He acknowledged the lessons without reservation, leading him to coin the term "Gurudev" for Tagore, which means "venerable teacher." A modest Tagore did not aspire to be a national role model, and he was quick to publicly state that he could never have given India the political leadership that Gandhi provided. In fact, it was he who bestowed on Gandhi the immortal epithet of "Mahatma" or the "Great Soul," which became synonymous with the man who came to be regarded as the father of the Indian nation.

Fundamental Elements for Studying This Conflict-Ridden Relationship

The clash between Tagore and Gandhi was striking and almost inevitable. Tagore was the poet, Gandhi the political crusader; Tagore believed in art, Gandhi in action; Tagore celebrated fullness of life, Gandhi renounced it; for Tagore the ultimate quest was beauty, for Gandhi it was truth. The range of differences, therefore, on a wide array of topics went beyond the noncooperation movement and the *Charaka* (spinning wheel) issue, which most critics regularly deliberate about. The poet and the Mahatma were conditioned by differing attitudes and approaches and often found themselves exchanging arguments assailing each other's position. Although the attacks remained at the intellectual level, the very process of being questioned was essential for Gandhi, as in the normal scheme of things there were few people who ever challenged his beliefs during the early years of India's freedom struggle.

Gandhi invented "the politician as a saint" image through his lifestyle and public conduct. The public's portrait of him was like that of Christ and his fol-lowers. But few are aware that Tagore, whose work and fame preceded Gandhi,

was seen as a carrier of exotic wisdom, often attributed to him for his physical appearance and demeanor—handsome, bearded, sage-like, quintessentially a mystical poet in the West. Tagore humbly accepted this image but sought to build his reputation on more solid accomplishments.

Tagore was not a politician, nor was he interested in wielding power over the lives of others, for good or for evil. But he had a clear and steady vision of man's destiny and unerring instinct. Tagore had no illusions about what is called progress in Western countries, which was at that time synonymous with luxuries and mechanized living. By progress he meant a continuous development of the human personality, both individual and corporate: "I believe in life only when it is progressive and in progress only when it is in harmony with life" (Tagore 1917). Gandhi found these qualities in Tagore worthy of emulation. But where the two figures differed was in their attitudes toward personal life. Tagore was more practical and worldly, advocating family planning instead of Gandhi's total abstinence. Hence, while Tagore's take on this issue appealed to the common man, many could not comprehend how Gandhi could, on one instance, advocate celibacy to a newly married couple on his Phoenix farm in South Africa, after he himself had fathered four sons.

That Tagore was intellectually more perceptive than Gandhi, and fair-minded at that, comes across more explicitly when compared with Gandhi's conservative instincts regarding the country's past traditions. Gandhi ignored reason and logic, depending more on spiritual instincts, often bordering on blind belief. Tagore, then, did not hesitate to blame Gandhi for exploiting this irrational force, which prevented the masses from progressing with the rest of the world. It was no surprise, therefore, that within a few years after Gandhi's death a large number of educated Indians began looking to Western liberalism for advancement rather than looking within the Indian system of beliefs.

This kind of blind faith also pervaded Gandhi's views on nationalism. Tagore had the greatest admiration for Mahatma Gandhi as a person and as a political leader, but he was also highly skeptical of Gandhi's form of nationalism. Consequently, guided by their respective life philosophies, each sought to express his ideas regarding India's freedom. Thus while Gandhi opined that India should be free in order to serve humanity and do it more actively and purposefully, Tagore believed that India should serve humanity in order to be free.

The basic difference in approaches was to manifest itself in each of their positions on internationalism. It was natural for Gandhi to insist on national organization before embarking on any form of internationalism, as he insisted on being a nationalist before becoming an internationalist. Tagore's view was diametrically opposite, as he held that a nation did not have to become strong and independent before raising its voice in the world. He argued that India, in spite of her penury and humiliation, could still come to the rescue of humanity. In fact, Tagore saw no meaning in political freedom when the mind was not free.

Tagore Wins the Nobel Prize for Literature

On November 13, 1913, the Swedish Academy in Stockholm announced that Rabindranath Tagore had won the Nobel Prize for Literature. This was largely for his book *Gitanjali*, which was published in English, in a limited edition, and also in a public edition. Tagore was the first Asian to be awarded the prize, and the media were surprised at the announcement of the award, partly because many had not heard of him.

The first public comments on the announcement were made by the Swedish writer Per Hallström and the poets Erik Axel Karlfelt (who went on to win the prize in 1931) and Verner von Heidenstam (who won the prize in 1916). They commented that although Tagore's books "have only recently become known in the Western World, [they] show an original and poetical vein of great depth and undoubted literary merit." Hallström later became chairman of the Nobel Committee of the Swedish Academy for the Nobel Prize for Literature, a post he held from 1922 until 1946.

Gandhi and India in later years were to endorse this very thought when Tagore's haunting lines from his *Gitanjali* were sung at several national gatherings:

> *Where the mind is without fear*
> *And the head is held high . . .*
> *Into that heaven of freedom,*
> *My Father, let my country awake.* (Tagore 1924: 28)

Tagore's reservations about patriotism were based on areas where India could have limited one's freedom to engage in ideas from outside the "narrow domestic walls" and failed to support the causes of people in other countries. From the time he received his Nobel Prize in 1913, Tagore stressed his international concerns and even denounced the excesses of nationalism not only in India but also in Japan, China, and the West. He went on to assert that contact with the West, especially with Britain, produced a kind of energy and a heat of enthusiasm in the minds of the Indian people that had lain dormant and inactive for some time. India's connections with the British, he maintained, helped her people renew their faith in their religion, in their philosophy and culture, and rediscover their country in a new spirit of hope and aspiration. Gandhi might have accepted the efficacy of Tagore's statement had he lived longer and seen the realization of his belief. It was unfortunate that Gandhi could not witness the international impact that his own philosophy would have, particularly on the civil rights movement of Martin Luther King, Jr., and Nelson Mandela's struggle against apartheid in South Africa.

Tagore's treatment of issues and policies was mellow and masterful, the result of a long intellectual introspection on the land, its people, and the past. Gandhi's

reaction to issues was often impulsive. Seeking reason and in a quest for answers, Tagore posed a series of questions to Gandhi's curious statements. This was something Gandhi only comprehended later when he attempted to understand Tagore.

Tagore questioned the viability and sustenance of several of Gandhi's personal beliefs. Although he himself shared several of these, he differed from the extremity of Gandhi, which in the long run made the poet's views seem more practicable. On asceticism, Tagore accepted the ideal of simplicity and self-restraint, but he refused to glorify it. He seemed right when he asked Gandhi if the extreme lengths of the sacrifices he prescribed took into account the limitations of human nature. On suffering, Tagore believed that this was as much an aspect of life as pleasure was. Therefore, was Gandhi not overlooking the Vedic emphasis of *ananda* or bliss by overestimating the importance of suffering?

Regarding patriotism, Gandhi sought instant patriotic fervor from Indians who historically had submitted themselves to centuries of foreign rule. For Gandhi, patriotism was the same as humanity. Tagore, on the other hand, believed that the complete man must never be sacrificed to the patriotic man, that humanity was rich, large, and multifaceted. His criticism of patriotism was a persistent theme in all his writings.

It appeared as if Gandhi had cashed in on *blind obedience* when he appealed to the momentary passion of youth when he called for a boycott of schools and the purchase of foreign cloth to register protest against the British rule. He even termed it a sacrifice. Tagore thought this to be shortsighted and that it dragged unconvinced students into the muddy waters of politics. He then questioned the exact nature of the sacrifice that the students were being asked to make. When Gandhi suggested that every patriotic Indian must spin cotton yarn as a religious duty, Tagore questioned the logic behind it, as it seemed as if Gandhi believed that spinning was the only panacea for all the ills of India. What was most
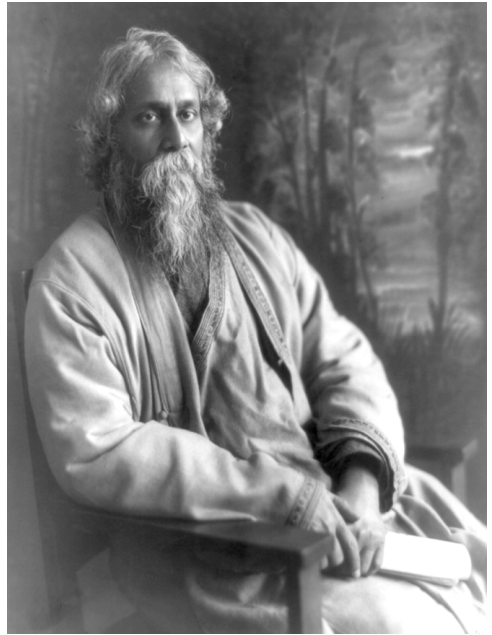
Rabindranath Tagore ranks among history's most renowned poets. His prodigious literary output and support for the cause of Indian independence made him a popular writer in his native Bengal and other provinces of British India in the late-19th and early-20th centuries. His international fame was sealed with a Nobel Prize in literature, awarded for his work *Gitanjali* in 1917. (Library of Congress)

unacceptable to Tagore was the fact that in the name of the nation and in the name of Gandhi as the national leader, absolute obedience and conformity were being demanded of the unsuspecting masses.

On the sin of the debilitating caste system in India, Tagore in the 1930s was more concerned about its divisiveness, both in Bengal and India as a whole. With his acute sensitivity and worldwide experience, he perceived its insidiousness with greater clarity than any other Indian, including Gandhi. When a terrible earthquake struck the northeastern state of Bihar in 1934, Gandhi attributed it to God's punishment for the Indian peoples' practice of untouchability. Tagore found that Gandhi was exploiting a terrible human tragedy, playing upon the fears of divine retribution for human sins, in support of a cause he was espousing. Trying to reason with Gandhi, he pointed out that the Mahatma was violating his own principle that the means should be as pure as the end.

On Gandhi's spinning as the means of attaining *Swaraj* ("self-rule"), Tagore felt that the call to self-rule was restricted to one narrow field alone. He wondered if this could be the new age of a new creation. Tagore put forth a more enlightened argument that even though there was plenty of home-spun thread during the Mughal invasion, this thread did not bind India together or prevent her from completely falling apart. When Gandhi insisted on the charka being an economic salvation for India, Tagore questioned what economic or mental profit could result from merely spinning.

On cooperation itself, Gandhi stressed the need for Indian self-sufficiency in every sphere before cooperation with the rest of the world. But Tagore saw the need for international cooperation and placed India in a position to deliver a message to the world and incorporate others messages in her own cultural repertoire.

The argument over English as the medium of education brought strange responses from Gandhi, who went to the extent of stating that boys studied English to get white-collar jobs, and girls to procure husbands with enviable economic and social status. However repugnant the thought may have been to Gandhi, young Indians have studied English long after the British departed from India, precisely for the reasons he had stated. Tagore has been proven right once again, for the world today looks to India as a reservoir of human resources communicating effectively through English.

Further, Tagore made a point never registered before—the West had misunderstood the East, which was the root of all disharmony between them. In order to set that right, the East could not reciprocate in the same manner. His age, the poet felt, was possessed by the West, made possible by the kind of mission for mankind that she had. Therefore, it was necessary for the East to go to her and learn whatever she could. A time would come, he believed, when the West would come to the East to learn from her and make it their home. His statement seems so prophetic of the present times where India occupies such a focal point of connectivity in a globalized world.

Tagore even appears like a soothsayer of our times when one studies the experiment in global education that he undertook at his own world university in Shantiniketan and the programs at Sriniketan and Visvabharti universities. Truly great was his influence on Gandhi in this sphere, for the Mahatma was to replicate many of Tagore's principles in his educational plans.

Tagore's persistent questioning seems to have made a stance-changing impact on Gandhi. This was evident when the poet argued with Gandhi when he erroneously called Indian social reformers Raja Ram Mohan Roy and Lokmanya Tilak "pigmies" when compared with other Indian greats, as they had advocated thinking and writing in the English language. Openly confronting Gandhi, Tagore managed to get him to look beyond his narrow thinking and even squeezed out of him that argument-clinching *mantra* so often quoted in today's multicultural times: "I want the cultures of all the lands to be blown about my house as freely as possible. But I refuse to be blown off my feet by any" (Gandhi 1968: vii).

Historical Proofs on Tagore's Role in Gandhi's Greatness

History bears witness to Tagore the trailblazer who established institutions and policies, many of which were consolidated by Gandhi later and even won him acclaim. Long before the historic Swadeshi movement of 1920, Tagore had begun to propagate the doctrine by opening a Swadeshi Store in Calcutta where he had assembled for sale from all parts of the country various articles of Indian manufacture required for daily life. Tagore even created a design for a typical "national" costume for the whole of India, as part of the Hindu Mela, to encourage the indigenous arts.

Tagore had experimented with various educational prototypes, beginning as early as 1901 when he established Shantiniketan. Rich new ideas were amassed during his 40 years as a school master in humble village surroundings. Tagore commenced experiments with what came to be known later as "community development" programs. He in fact produced the first manual on rural education and took a leading part in establishing the Indian National Council of Education in 1916. Many of his suggestions were accepted later by Gandhi. Gandhi even employed his first batch of teachers for his Basic School from Shantiniketan. Many of Tagore's initiatives between1892–1910, though unheeded at that time, later became the foundation for Gandhian educational initiatives.

Few people know that even before Gandhi had advocated and applied nonviolence, Tagore had hailed the advent of such an apostle in his drama *Prayoschitta* (Atonement), publicized in 1909, based on a novel written as early as 1883. The character in the play took on the leadership of his unarmed people in a nontax campaign against cruel eviction. All the elements of a regular Gandhian *satyagraha* based on truth, nonviolence, and fearlessness were depicted in this play, appearing to be a prelude to the actual Indian National movement.

Gandhi arrived in India in 1914, and it took a while for him to identify a role for himself in the country, even though he had sown the seeds of an uprising in South Africa. But Tagore had stolen the march from him on many fronts between 1914–1916. In 1915, Tagore's relationship with British officials was considerably warmer than that with Gandhi. Nine days after Gandhi's Phoenix party left Shantiniketan, Tagore received the first governor of Bengal, Lord Carmichael, and in June that year the poet was knighted. For a few years until the Amritsar massacre of 1919, Tagore was able to influence the work of the councils of the Raj. But this was not through pandering to the British but through a justified criticism of their negative actions. He also renounced his knighthood soon after the Amritsar incident, months before Gandhi followed suit.

Tagore's world recognition in the early part of the 20th century made it that much easier for Gandhi to become a world leader. Critics would agree that Tagore's Nobel Prize for Literature, in a sense, put India on the mental map of the world. It carved a niche for Indian presence in England, in particular, for in his many essays and articles on Indo-British relationship, such as "East & the West" and "The Great & the Small," Tagore had paid generous tribute to the greater minds of Europe and England in particular. Tagore's repeated foreign tours attracted a stream of admirers of Europe to Shantiniketan. Among them were C. F. Andrews, W. W. Pearson, Romain Rolland, Erza Pound, W. B. Yeats, Earnest Rhys, and Albert Einstein. And it was at the initiative of the painter William Rothenstein that Tagore's *Gitanjali* became known to the West. Gandhi did learn from Tagore's Japan tour, where the poet appreciated the awakening of the Japanese people under Western influence and admired them for retaining the spiritual tradition of the East. In comparison, Gandhi's utilization of Western influence was very personal and not to the advantage of the greater numbers. Gandhi had lived with Englishmen both in England and South Africa for nearly 30 years, yet his policy was often confrontational as he believed that one had to stand up to an Englishman in order to gain his respect.

The most contentious issue in the Gandhi–Tagore debate, nonviolent noncooperation, would prove in the end that Tagore's predictions would come true. Tagore felt that Gandhi's call for open noncooperation had not been thought through clearly. He read a paper called "The Call of Truth" in 1921 at the Calcutta University Institute, in which he openly criticized Gandhi for exploiting the public sentiment as well as their excitement as a means of ensuring some quick political gain. He positively condemned the movement as a symbol of *himsa* or violence, as he thought it rested on the hatred of others and on a negative foundation. The poet also bitterly condemned the growing mood of lawlessness and indiscipline among the public. He argued that while millions were dying in the cold, it was inhuman to amuse oneself by burning clothes. According to the poet, cultural seclusion from the West was not good, as India could receive the best in the Western culture as an "awakening of India is a part of the awakening of the world." He added, "from now onwards, any nation which takes an isolated view

of its own country will run counter to the spirit of the New Age and know not peace" (Tagore 2007: 545-565).

Noting the abrupt end to the noncooperation movement, one is automatically drawn to the example Tagore had cited in support of cooperation. He cited the success of Horace Plunkett's application of this principle in the economic reconstruction of Ireland. Tagore saw the benefit in applying the Irish experiment to India, but Gandhi merely dismissed the suggestion, stating that the Indian circumstances were different.

Over a period of time, however, Gandhi realized the truth in Tagore's arguments, especially as they pointed to the futility of the Mahatma's extreme measures. Some of this could be attributed to the poet's quiet involvement in the political aspects of the freedom struggle since the early 1930s. Tagore began making bold statements on the British-Indian relationship, which according to his remarks at a London Quaker meeting in 1931, was characterized by "the dark chasm of aloofness." His public tone brought Gandhi closer to the poet, after a decade of virtual estrangement. Though he continued to disagree with Gandhi periodically in private, there were relatively fewer open conflicts between them in the last 10 years of Tagore's life.

It is interesting to note how much Gandhi counted on the valuable opinion of his "Gurudev" during times of mental stress and trial. It was poignant to observe Tagore's presence at Gandhi's bedside at the Yaravada jail in Poona in September 1932, when the Mahatma broke his epic fast after protesting against the British prime minister's verdict on the Communal Award. Gandhi himself would write later that Gurudev was the richest find among the undreamt of treasures that the fast brought for him.

Scholarly interpretations of Tagore's legacy highlight the poet's indelible mark on his times, and subsequently on Gandhi. Leonard Elmhirst, the English philanthropist and social reformer who had worked closely with Tagore on rural reconstruction in India, found the experience brought together a resource base for Gandhi. Literary luminaries Ezra Pound and W. B. Yeats located in Tagore's literary framework a whole people and a whole civilization that were immeasurably strange to them, yet they reflected their own images in his expressions. Even ardent Gandhi admirers Albert Einstein and Romain Rolland marveled at Tagore's innate ability to influence the freedom movement as an artist, without getting lost in literary sentimentality. Historian H. G. Wells remained enchanted by Tagore's unique explanation of how the 19th century had made the West feel racially superior, but that the East would manage to turn the tide by assimilating the sciences.

Fundamentally Tagore was humble, ever willing to learn from constructive criticism. This is where Gandhi appeared in a lesser light, particularly with regard to opinions on nationalism. In Dennis Dalton's (1993) view, Gandhi an unremitting nationalist, not merely endorsed nationalism but also dismissed all attacks on it. Tagore, while critical of the cult of nationalism and the unthinking

sacrifices it demanded of the youth of India, was quite sympathetic to nationalism insofar as it was congruent with both "internationalism" and the ethos of Indian civilization. According to sociologist Ashis Nandy, a fear of nationalism was common to both Tagore and Gandhi. He believes that Gandhi and Tagore did not want their society to be caught in a situation where the idea of the Indian nation would supersede that of the Indian civilization.

The Gandhi–Tagore debate had the same kind of resonance as other controversies that would follow later: the Gandhi–Ambedkar, Gandhi–Nehru, and Gandhi–Godse debates. This debate was kept alive through the 1920s and 1930s by the poet's resolute opposition to charkha. Tagore simply pointed out that the charkha was being accorded a higher place than it was due. He criticized Gandhi in his 1925 article "The Cult of the Charka" (Tagore 2007: 730-744). Gandhi strangely went on to romanticize his belief by declaring that he would wander after his beloved Sita the charkha, and deliver her from the 10-headed monster from Japan, Manchester, and Paris (Gandhi 1956: 227-231). Although Gandhi's doctrine of noncooperation did not really set up a barrier between the East and the West, Tagore knew that it would be interpreted as doing so once Hindu nationalism was stirred. C. F. Andrews, the English missionary who was close to both the men, realized Tagore's prophetic warning when a frenzy of violence broke out unleashing rioters "like a bunch of swirling dervishes," as Tagore put it. Gandhi acknowledged his "Himalayan miscalculation" and fasted to atone for his mistake. A major point that Tagore scored with critics in this phase of the debate was when Gandhi prematurely declared noncooperation and promised self-rule within a year, only to immediately suspend it. The poet plainly was asking if the Hindus and the Muslims were supposed to just lie down peacefully after the British left India.

Writers have registered the fact that as more staunch and educated Indians came into the freedom fray in the 1930s, there appeared several cracks between them and Gandhi. The Gandhi–Nehru divide was a case in point. In contrast, Tagore's stance through all of these phases remained unchanged. His criticism of the British administration of India was consistently strong and grew more intense over the years. But his criticism did not mean a denigration of the Raj or Western culture. On the other hand, Gandhi's stand on several issues seemed disjointed, eliciting even unexpected remarks from him on various occasions. His quip in reply to a question asked in England as to what he thought of Western civilization was: "It would be a good idea." This was something that could never have come from Tagore's lips in spite of provocation.

Conclusion

A reasonable critic would accord an equally illustrious status to both of these stalwarts of the Indian freedom struggle. While Gandhi continues to shine brightly in the galaxy of the world's greatest, the path to that greatness was set by none other

than Tagore. The framework for a comparison of the two men becomes possible only because basically they were not very different from each other.

Tagore owed a great deal to his travels and encounters in the West for which he believed he was privileged. He believed that his life followed the course of his celestial namesake the sun and the last part of those honors he owed to the West. He respected and rejoiced in the sweet appreciation from fellow human beings who were so different from him. Tagore regretted India's segregation from the rest of the world, and his message was that India, in order to find herself, must give of herself. The fruit of his advice is being borne in full, in India of the 21st century.

There was no reconciliation between Tagore and Gandhi on issues on which they held firm views. Both possessed deep emotional intensity, as was obvious in the exchanges they had on a series of issues between 1921–1925. And the controversies had no secrecy about them. But their intellectual honesty made them respect each other all the more. As time progressed, each considered the other a giant, towering above any other living Indian. They soon exchanged worthy praises. Tagore said that the spirit of the Mahatma was infused into the whole of India, chasing away the faintness. Gandhi said of Tagore on his 70th birthday that the poet's genius and purity of life had raised India in the estimation of the world. Gandhi on this occasion contributed to *The Golden Book of Tagore*, a global anthology sponsored by him, Einstein, Rolland, and the Greek poet Kostes Palamas among other greats. Tagore's composition "Ekla chalore" ("If nobody responds to your call, move along") was Gandhi's favorite song. Gandhi and Nehru expressed appreciation of Tagore's stupendous role in the national struggle and as a fitting tribute chose his song "Jana Gana Mana Adhinayaka" ("The leader of people's minds") as independent India's national anthem. An added feather in Tagore's cap was that another composition of his "Amar Sonar Bangla" was chosen as the national anthem of the newly born nation of Bangladesh in 1971, proving the universal relevance of Tagore's verses. He thus becomes the only individual ever to have authored the national anthem for two different countries.

One has to be great in order to lead another to greatness. Tagore was great and had led the Mahatma to even greater heights of world renown. Bertrand Russell said it all when he wrote that what Tagore had done for Europe and America, by the way of softening prejudices and removing misconceptions, was worthy of the highest honor. Like the sun after which he was named (Rabi meant the sun), he shed light and warmth on his age, vitalized the mental and moral soil of his land, and revealed unknown horizons of thought. The Mahatma walked ahead led by the sun.

References and Further Reading

Ahluwalia, B. K., Ahluwalia, Shashi. *Tagore and Gandhi: The Tagore-Gandhi Controversy*. New Delhi: Pankaj Publication, 1981.

Dalton, Dennis. *Mahatma Gandhi: Non-Violent Power in Action*. New York: Columbia University Press, 1993.

Dutta, Krishna, and Andrew Robinson. *Rabindranath Tagore the Myriad-Minded Man* London: Bloomsburg Publishing, 1997.

Gandhi, Mahatma. *The Gandhi Reader: A Sourcebook of His Life and Writings*. Edited by Homer A. Jack. Bloomington: Indiana University Press, 1956.

Gandhi, Mahatma. *The Selected Works of Mahatma Gandhi*. Ahmedabad, India: Navajivan Publishing House, 1968.

Kripalani, K. R. *Gandhi, Tagore and Nehru*. Bombay: Hindu Kitab Limited, 1947.

Naravane, V. S. *The Saint and the Singer: Reflections on Gandhi and Tagore*. Allahabad: Publisher Naravane, 1995.

Prabhu, Kelekar Ravindra, ed. *Truth Called Them Differently, Tagore-Gandhi Controversy*. Ahmedabad: Navjivan Publishing House, 1961.

Rolland, Romain. *Mahátmá Gandhi: The Man who Became One with the Universal Being*. London: Allen and Unwin, 1924.

Roy-Choudhury, Sumitra. *The Gurudev and the Mahatma*. Pune: Shubhada-Sarasvati Publications Limited, 1982.

Tagore, Rabindranath. *Gitanjali: Song Offerings*. New York: Macmillan, 1924.

Tagore, Rabindranath. *The English Writings of Rabindranath Tagore, Volume 5*. New Delhi: Atlantic, 2007.

CON

Mahatma Gandhi's (October 2, 1869 to January 30, 1948) association with Rabindranath Tagore (May 7, 1861 to August 7, 1941) has been a theme of intense historical debate in shaping rich contours of modern Indian history, which led to the controversies and notion that Gandhi would not have been a world leader without the influence of Tagore. Moreover, the proposition that Tagore was essential to Gandhi was further strengthened because Gandhi addressed Tagore as "Gurudev" (Godly Preceptor) in his correspondence, meaning *teacher and guide* in Indian tradition.

It is no doubt that Gurudev Rabindranath Tagore, a Nobel laureate and illustrious painter, a play and story writer, novelist, poet, singer, educator, internationalist thinker, and politician, has many golden faces, has immensely influenced the sociocultural-political movement in modern India. But it would be utterly wrong to analyze Tagore's personality and thought as an essential element in creating Gandhi as a world leader. In fact, in reality, Tagore and Gandhi's thoughts of action are two different perspectives to view Indian sociopolitical problem under British colonialism, and both of them had different

programs of action to ameliorate the conditions of Indian people suffering under British imperialism. If we accept that Tagore was important in making Gandhi a great world personality, it would be a historically prejudiced notion, negating the role of the individual in history in attaining greatness. Greatness is formed through commitment of inner consciousness to work toward a cause developed from one's own experiments as best fits the circumstances. As well as historically, it is also unnatural to accept the argument that Gandhi was known to the world only from his Indian experiment with the truth, negating his important work carried forward in South Africa since 1893 until his entry into the Indian political scene in the 1917 Champaran movement in Bihar.

Mahatma Gandhi's addressing Tagore as "Gurudev" was more related to paying respect to a world famous poet and Nobel laureate writer rather than considering Rabindranath as a guide and teacher. A title of respect is given to those who contribute in a significant way to the society and was part of Gandhi's personality. On another occasion, one finds that Gandhi, who himself was known as Mahatma in later years, gave the title of Mahatma to Arya Samaj leader Swami Munshiram in his correspondence, later known as Swami Shraddhanand of Gurukul Kangri, Hardwar (a holy place in present-day Uttaranchal state of India) in March 1914, a year before coming to India. It is also important to note that Gandhi used the word Gurudev (first time in 1914) for Tagore as well as for Rabindranath Tagore in most of his 1920 correspondence with Tagore. Moreover, for Gandhi, Tagore had been always a poet par excellence from Bengal and not a politician, as he always referred to Tagore as a great poet, intellectual, and writer but never mentioned him as a politician in his writings, despite Tagore's active participation in the 1905 antipartition Swadeshi movement in Bengal and Gandhi's firm knowledge about that.

It was from Gopal Krishna Gokhale that Gandhi publicly acknowledged on many occasions that he learned the nuances of politics, and he was his teacher, friend, philosopher, guide, and political guru. Gandhi's firsthand detailed acquaintance with Tagore came through C. F. Andrews's lecture on the poet at City Hall, Cape Town, South Africa on February 12, 1914. He informed his political master Gopal Krishna Gokhale through a cable to Poona, India, on February 18, 1914, showing appreciation for Andrews's efforts in highlighting the works of Tagore and values of higher Indian life (Gandhi 1998: 71). Gandhi accepted this fact, and while speaking at his reception ceremony given by British and Indian friends attended by luminaries of Indian independence movement such as Sarojini Naidu, Sachidanand Sinha, Lala Lajpat Rai, and M. A. Jinnah, at Hotel Cecil in London on August 4, 1914, expressed appreciation for the efforts of C. F. Andrews, stating: "Mr. Andrews have no notion what he did; how he worked in selfless zeal, preaching love for India through his Master—the poet-saint at Bolpur [in present day West Bengal of India] whom I have come to know through Mr. Andrews— Rabindranath Tagore" (ANC website). It is also important to note that Andrews

---

"Mahatma" . . . What's in a Name?

The term *Mahatma* was used in India to describe a great and inspiring teacher and is a Sanskrit word for "Great Soul." In that context it was used by Rabindranath Tagore to describe Mohandas Gandhi soon after Tagore had won the Nobel Prize for Literature.

The term started to be used to describe people in theosophical literature in the 19th century, especially by Helena P. Blavatsky. It was often used to refer to people who had gained "higher" learning in northern India, especially Tibet, being used in the title of Edwin G. Scary's book *In Search of the Mahatmas of Tibet* (1937). It has also been used pejoratively, as in Edmund Garrett's *Isis Very Much Unveiled: The Story of the Great Mahatma Hoax* (1894).

The British newspaper *The Times* first used the term Mahatma to describe Gandhi in October 1920, and two months later, it remarked "his austere asceticism and other-worldliness have earned for him the name and reputation of a mahatma, i.e. of one on whom the mantle of wisdom of the ancient Rishis of the Vedic age has fallen. As such he is outside and above caste."

---

was first introduced to Gandhi through Gokhale and not by Tagore. Even after the Cape Town lecture by Andrews on Tagore, Gandhi wrote from Cape Town, South Africa, to Gokhale in India on February 27, 1914: "My present ambition you know. It is be by your side as your nurse and attendant. I want to have the real discipline of obeying someone whom I love and look up to." He expressed his desire to return to India and further said, "I'm entirely in your hands. I want to learn at your feet and gain necessary experience" (Gandhi 1995c: 360–61).

Moreover, before his final return to India on January 9, 1915, Gandhi had already secured the Indian Relief Bill from the South African government and British Empire in 1914 for indentured Indians and introduced his most powerful political weapon "Satyagraha" (Sat: Truth; Agraha: firmness, later also to be utilized in his Indian struggle) based on truth and nonviolence with firmness in 1906. He had practiced it in July 1907 when a resident Indian of Germiston (South Africa), Pandit Rama Sundara, was arrested, which made Sundara the first Satyagrahi prisoner in history and famous all over South Africa at that time (Gandhi 2003: 102, 128). In addition, before his arrival to India in 1915, Gandhi had been interviewed many times by the world press and news agencies such as Reuters and South African print media such as *The Natal Mercury*, *Rand Daily Mail*, *The Transvaal Leader*, *The Star*, *Pretoria News*, *Cape Times*, and *The Cape Argus* and was known by name in world politics, apart from his own weekly media venture in South Africa titled *Indian Opinion* (1903–1961). Gandhi had already conceived the idea of Indian home rule as a soul force before meeting Tagore and written a book titled *Hind Swaraj* ("Indian Home Rule") in November 1909, while returning back to South Africa on a ship, the SS *Kildonan Castle*,

which was published first in his weekly *Indian Opinion* prior to Annie Besant's and Bal Gangadhar Tilak's Home Rule League movements in India in 1915 and 1916, respectively (Gandhi 2003: 211).

Gandhi's first visit to Shantiniketan (a university founded by Tagore in 1901) with his Phoenix ashram teammates after arriving in India was on February 17, 1915, one month after his arrival at Bombay and was attended by his friend C. F. Andrews, as Tagore was out of station. His first face-to-face contact with Tagore was on March 5, 1915. Within two days of his arrival, Gandhi received the news of the death of Gokhale so he had to leave for Poona after two days of his stay at Shantiniketan to pay his last respects to his political guru. His visit to Shantiniketan was to learn more intimately about the poet whom he had heard much about from Andrews and to gain knowledge about the learning institution of Shantiniketan, which could be helpful in establishing his own ashram. After Shantiniketan he also visited the Gurukul of Swami Munshiram at Hardwar on April 6, 1915, as well as Mohini Ashram and Ramakrishna Mission when visiting Rishikesh before establishing his own Sabarmati Ashram or Satyagraha Ashram near Ahmadabad on May 25, 1915 (Gandhi 2001: 329). During this visit to Shantiniketan, Gandhi strictly advised his teammates to scrupulously observe his own Phoenix Ashram rules of South Africa, which itself implies following one's own ideas rather than those of Shantiniketan. It even suggested that Shantiniketan staff and Tagore dispense with the services of paid cooks and cook their own food themselves, which was gladly accepted by the poet, and he also suggested that this experiment contained the "key to Swaraj" ("self-independence") (Gandhi 2001: 317–18). One important lesson he learned from his stay at Shantiniketan was that "the Scavenger's work would be our special function in India" (Gandhi 2001: 324).

Even after his visit to Shantiniketan and meeting with Gurudev in 1915, Gandhi preferred to give instructions, in March 1921, to fellow countrymen at Berhampur Municipal Corporation (in the present-day state of Orissa in India) at a reception for him, to follow the noble traditions of Gokhale and not of Tagore; and he rebuked the members of the organizing committee for welcoming him in English language and said that "Gokhale's supreme work consisted in spiritualizing the politics of the country and national form of education was not complete without the spinning-wheel and Hindustani" (Gandhi 1994: 484).

Gandhi, who had traveled widely to the West before beginning his political career in South Africa and India, was meticulous and farsighted in his planning of Sabarmati Ashram, and his ashram list of goods included the maps of India, maps of the Bombay presidency, map of Gujarat, and, most important, the map of the world, besides utensils and other household utilities, to embark on the final journey as a great leader against British colonialism developed from his own experiences of life and experiments with truth rather than on Tagore's philosophy and help (Gandhi 1994: 86–87). Even Gandhi's first experiment with

truth in India during the Champaran movement (in Bihar, India) in 1917 was based on his own consent and thought devised accordingly and locally as the movement progressed to success without seeking help from Tagore. It is here that he first realized face-to-face contact with God, *ahimsa* (nonviolence), and truth, while meeting peasants protesting against land revenue policy of the government. He always believed more in his own experiences of life, so he decided in consultation with his coworkers at Champaran that nothing should be done in the name of the congress, which was a practical political strategy as congress was unknown in the locality (Gandhi 1994: 343–44). Even in his great work on his experiments with truth at South Africa titled *Satyagraha in South Africa*, first published in 1928, he does not mention Tagore, further proving that the poet was not essential to Gandhi, although Gandhi could have accommodated Tagore's thoughts and works as an inspiration for the book he published when he was in India after learning about Tagore's greatness.

Gandhi was convinced that because Satyagraha was successful in South Africa it should be tried in India. "I have no doubt," he said, "that the British Government is a powerful Government, but I have no doubt also that Satyagraha is a sovereign remedy," which he later successfully experimented with at Champaran in Bihar (1917) and in Ahmedabad and Kaira in Gujarat in 1918 (Chandra et al. 2001: 122).

Gandhi and Tagore: Two Sides of a Coin of the Independence Movement in India in Their Journey toward the International Stage

Gandhi and Tagore had different perspectives toward India's amelioration against British imperialism, and these could be found much before the first major nationwide Satyagraha in India against the Rowlett Act in March 1919 on the question of vernacular language versus English language. In January 1918, when Gandhi was seeking opinions of Indian leaders on this issue to formulate a policy to promote Hindi language as the only possible language for interprovincial intercourse and for all other national proceedings, Tagore wrote back to Gandhi that "of course Hindi is the only possible national language for inter-provincial intercourse in India. But . . . I think we cannot enforce it for a long time to come" (Tagore 1996: 736). And after the noncooperation movement (January 1921–February 1922), it was endorsed in the September 1920 Calcutta Congress Special Session and December 1920 Nagpur Congress Session against British rule, and the difference of opinion widened further when Tagore criticized noncooperation programs, including promotion of Hindi. While replying to Tagore's criticism Gandhi wrote:

> The Poet does not know perhaps that English is today studied because of its commercial and so-called political value. Our boys think, and rightly in the present circumstances, that without English they cannot get Government service. Girls are taught English as a passport to marriage. . . . I cannot

tolerate the idea of parents writing to their children, or husbands writing to their wives, not in their own vernaculars but in English. I hope I am as great a believer in free air as the great Poet. I do not want my house to be walled in on all sides and my windows to be stuffed. I want the cultures of all the lands to be blown about my house as freely as possible. But I refuse to be blown off my feet by any. I refuse to live in other people's houses as an interloper, a beggar or a slave. (Gandhi 1968: vii)

Gandhi wanted the Hindi language to serve as a bridge to different languages and cultures of India and said that major works of other languages should be translated into it, differing with Tagore's proposition of continuing with English language until an appropriate time arrived. In a July 9, 1938, article in *Harijan* (started in 1933, English weekly published under the auspices of Harijan Sevak Sangh and supervised by Gandhiji) on "Higher Education," he further advocated for vernacular languages to build India on its own heritage and says:

I must not be understood to decry English or its noble literature. The columns of "Harijan" are sufficient evidence of my love of English. But the nobility of its literature cannot avail the Indian nation any more than the temperate climate or the scenery of England can avail her. India has to flourish in her own climate and scenery and her own literature, even though all the three may be inferior to the English climate, scenery and literature. We and our children must build on our own heritage. (Gandhi 1994: 161)

Moreover, Gandhi was smart enough to understand the criticism of a personality like Tagore to the noncooperation movement and its political ramifications, so he was always prompt in replying to every questions raised by the poet, not forgetting to mention him as a "great Poet" or even recalling Lord Harding's given title to Tagore "The Poet of Asia" or as "Dr. Tagore," not recognizing him as a politician or even Gurudev. When writing to his friend and disciple of Gurudev C. F. Andrews on March 28, 1921, from Puri, whom he used to address "Charlie" with love, he wrote that "I'm unable to agree with the implications of Gurudev's letter reproduced in the 'Tribune' and sent to me by a friend for reply in 'Young India.' I have glanced through it once and I could not help thinking that he had not understood the simple beauty and the duty of non-cooperation" (Gandhi 1994: 483).

However, he promptly replied to the poet's criticism of the noncooperation program of boycotting government schools and said:

Nor need the Poet fear that, non-cooperation is intended to erect a Chinese wall between India and the West. On the contrary, non-cooperation is intended to pave the way to real, honorable and voluntary co-operation based on mutual respect and trust. The present struggle is being waged against compulsory co-operation, against one-sided combination, against the armed

imposition of modern methods of exploitation masquerading under the name of civilization. Non-cooperation is a protest against an unwitting and unwilling participation in evil . . .

Here I must differ from him. I have never been able to make a fetish of literary training. My experiences has proved to my satisfaction that literary training by itself adds not an inch to one's moral height and that character-building is independent of literary training. I'm firmly of opinion that the Government schools have unmanned us, rendered us helpless and godless . . .

He has a horror of everything negative. His whole soul seems to rebel against the negative commandments of religion. . . .

I therefore think that the Poet has been unnecessarily alarmed at the negative aspect of non-cooperation. (Gandhi 1956: 219-221)

Tagore also objected to the burning of foreign clothes when a large number of his countrymen had to go seminaked; he could also not bring himself to believe that the spinning wheel was the panacea for all the economic ills from which the country was suffering. But Gandhi justified the bonfire of foreign clothes on the grounds that their use was sinful in a two-fold way. It was sinful to buy and use them in as much as they were the product of sweat labor. For him, it was also sinful because it meant so much work was taken away from our own spinners and weavers. He also replied vehemently to the poet's argument that, instead of burning foreign clothes, it would have been better to have given them away to poor people to hide their nakedness as "I must refuse to insult the naked by giving them clothes that they do not need, instead of giving them work which they sorely need" (Suda 1963: 296–97).

Gandhi was a very farsighted practical statesman to understand the changes taking place in world history due to World War I, and in his journey as a world leader, he saw a great opportunity in supporting the Khilafat questions of the Ali brothers (Mahomed Ali and Shaukat Ali) in 1919–1920 to achieve Hindu-Muslim unity, despite knowing that through it, it would establish an old dying institution in Turkey as opposed to embracing the changes of time, all without taking any advice from Tagore. As he writes to Khilafat leader Mahomed Ali on June 29, 1919, "I need hardly say that you are never out of my mind, although we may not for some time come to meet face to face. I am in close touch with our public men, as also with those in authority regarding the *Moslem question*" (Gandhi 1994: 404). By supporting it, he became the first Hindu politician in modern India to attract the largest support base ever achieved by any Indian of India's significant Muslim community. However, interestingly, his Satyagraha pledge taken on April 7, 1919, to observe civil disobedience against the Rowlett Bill included distribution of Mustafa Kamal Pasha's writings (Turkey's nationalist leader working

against Khalifat) and the prohibited literature included *The Life and Address of Mustafa Kamal Pasha*, apart from *Hind Swaraj*, *Sarvodaya or Universal Dawn*, and *The Story of Satyagrahi*, written by himself (Bandhu 2003: 76).

The difference that existed between Tagore and Gandhi was acknowledged by Tagore himself to Romain Rolland in June 1926. During his visit to Rolland in Villeneuve, Switzerland, June 21–29, 1926, Tagore speaks of his differences of thought with Mahatma Gandhi, and eminent writer Rolland captures the thought of the poet in his diary notes and writes, "He shows [poet] that in supporting the Indian Muslims as he did in the Khilafat affair Gandhi was not working, as he hoped, for the unity of India, but for the pride and force of Islam, factors which are at present emerging in violent Hindu-Muslim disturbances of which latter, cunningly supported by the British Government, are the instigators." He further writes that "Gandhi avoided a direct reply (as he often does) to the question of foreign cloth as an impure objects, and he believed in idolatry for the people of India. To which Tagore replies that this means he believes that the Indian people need lies, and if the people need lies, they have no right to liberty" (Rolland 1990: 61–62).

On the question of caste differences and untouchability, according to Tagore, "Gandhi has never sought to breach caste divisions [this is another of the points for which Tagore finds it hard to forgive him]. . . . No doubt for Gandhi this is a question of choosing the right moment. At present Gandhi wants to solve the political problem, which is relatively easy. He does not want to tackle the infinitely more complex and dangerous social problem yet" (Rolland 1990: 63). But for Tagore, how can one solve the political problem without first solving the social problem? It is like building one's house on sand. Tagore felt disappointed when Gandhi actively entered into Indian politics. He deplored that the overflowing wealth of Gandhi's love and faith should be made to serve political ends. Rabindranath felt exceedingly sorry that a precious treasure like the Mahatma "should be cast out on the frail barque of politics and subjected to the incessant lashing of the waves of conflicting and irritated passions" and regretted that Gandhi had wasted his spiritual resources upon political ends (Rolland 1990: 295).

Mahatma's difference with Tagore became a great public controversy over Gandhi's preference for instinctive faith over the claims of scientific inquiry when he pronounced that the devastating earthquake in Bihar in 1934 was a "divine chastisement" for the sin of untouchability. And Rabindranath Tagore reacted very strongly and criticized Gandhi not only for implying that God, in inflicting punishment upon sinners, was unable to distinguish between the guilty and the innocent, since an earthquake is indiscriminate in its destruction, but also for strengthening the forces of unreason which fostered the belief that cosmic phenomenon had something to do with the fate of human beings on Earth. However, Gandhi stuck to his position with his characteristic firmness and refused to entertain questions about the rationality of divine action and replied that "I'm not affected by posers such as 'why punishment for an age-old sin' or 'why punishment to Bihar and not

to the South' or 'why an earthquake and not some other form of punishment.' My answer is: I am not God. Therefore I have but a limited knowledge of his purpose" (Raghuramaraju 2006: 88–89). Gandhi's answer of divine judgment to earthquake for the sin of untouchability was guided, seeing a vast populace of India believed more in divine faith and judgment of God rather than scientific inquiry, which might have prompted him as a practical statesman to utilize this great opportunity to strike hard in removing social evil from the Indian way of life through faith rather than going with Tagore's vision of reasoning. It is important to note that a year before the Bihar earthquake in 1934 he had started his weekly English magazine *Harijan* to raise the issue of untouchability and the depressed class. And two years prior to that, in August 1932, he had faced the political challenge and crisis of the communal award to that depressed class intended to divide the Indian community by Ramsay MacDonald and the British Empire, as well as the emergence and challenge of new leadership of downtrodden Indian society of B. R. Ambedkar, which ultimately resulted in compromise through the Poona pact of September 24, 1932, with Ambedkar.

For Tagore, India's real problem is not a political one but rather a social one. An avid reader once asked Gandhi in November 1920: "Dr. Tagore says in his book on 'Nationalism' 'no enduring political freedom can be built in this country on the quicksands of social slavery' . . . and I know that you too were of this opinion some time ago," to which Gandhi replied, defending his new position of political action differing with Tagore, that, "I adhere to my statement that amelioration of radical social evils meant an attainment of swaraj, but I did not then realize that the British Government was the greatest of all the social evils with which society was cursed" (Gandhi 1994: 486–87).

It is not to Tagore but rather to American writer Henry David Thoreau and Russian writer Leo Tolstoy to whom Gandhi acknowledges that he based some of his writings and to whom he owes some of his thought, such as from *Civil Disobedience*. However, in actuality the thought developed by Thoreau was taken from an English translation of the "Upanishads" that had found its way into the Harvard library in Cambridge, Massachusetts, in the 1820s and was read by the young American writer. In it he found a concept of human dignity and endeavor that he gradually evolved into a theory of nonviolent social and political action. In the 1850s Thoreau, who demonstrated his belief in his own doctrine by going to jail in protest against what he considered an unjust law, spelled out his new concept in a pamphlet he titled "Civil Disobedience." Eventually this pamphlet found its way into the hands of Leo Tolstoy in Russia, who sent a copy to Gandhi in South Africa (Bowles 1969: 172–73). So the ideology that started from one place reached into the hands of a worthy torchbearer of Indian tradition and culture who began to practice it with his experiments with truth and who is always modest to acknowledge that "I have nothing new to teach the world. Truth and non-violence are as old as the hills" (Gandhi 1948: 13).
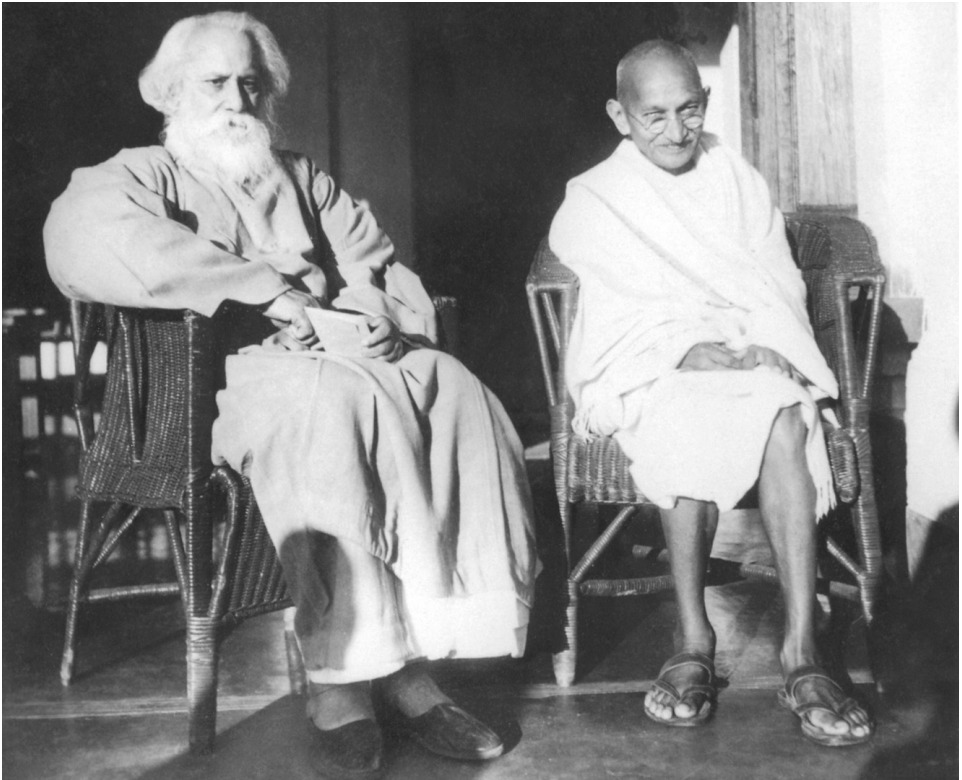
Rabindranath Tagore and Mahatma Gandhi at Santiniketan (West Bengal), February 18, 1940. (DPA/The Image Works)

More important, if Tagore had been essential to Gandhi's success as a world leader, then other imminent leaders of the world such as Martin Luther King, Jr., Nelson Mandela, and Daw Aung San Suu Kyi would have also recognized Tagore together with Gandhi in guiding their political actions.

However, it was Gandhian philosophy in Martin Luther's vision that captured the imagination of an entire generation of Americans, particularly young Americans, who gladly found a path to relinquish their prejudices and inhibitions. When Gandhi suggested many years ago that it may be the blacks who deliver "the unadulterated message of non-violence to the world," he demonstrated once again his extraordinary insight into the mind of a true world leader that evolved through his experiences of life (Bowles 1969: 171).

How much Gandhi and Tagore's personalities differed from each other has been best described by Romain Rolland in his diary notes of June 1926:

Tagore and Gandhi, two races of men and classes, one the aristocratic, the prince and other the popular Guru, the duel being between the prophet of religious and political action who scorns and debase intellectual values before

the divine word and moral values—and the supreme Artist, who lives in the firmament of his intellectual dream world. (Bowles 1969: 65)

However, these differences do not dispute the fact that Gandhi lacked any regard for Gurudev, which would be unnatural to his personality and philosophy to undermine and disrespect even enemies of his thought of action. Gandhi always wanted to place Tagore as a great poet and not as his political master and guide in his journey toward world leader. Moreover, it is interesting to note that it was not Tagore who was essential to Gandhi, but Tagore considered himself a disciple of Gandhi and wrote a poem titled "Aamraa Gandhi Mahaarajair Shishya" ("We are disciples of Mahatma Gandhi") when the Mahatma was visiting Shantiniketan to see ailing Gurudev in 1941, a few months before his death (Prasad and Tagore 2001: xiv).

Despite many differences on thought of action, Gandhi's reverence to Gurudev could be assessed from the fact that for a while he also considered accepting the presidency of Visvabharati to carry forward the works of the poet one month after Gurudev's death. As he wrote to Rathindranath Tagore on September 8, 1941, "If you and others want me to be President of Visvabharati, I must discuss the project with you. My whole soul is with you all in wishing to keep the triple institution (Visvabharati, Santiniketan and Sriniketan) fully worthy of Gurudev" (Gandhi 1998a: 303). However, due to his political engagement, constructive program, and old age, he refused to take the presidency of Visvabharati, and while congratulating the new president Abhanindra, he wrote to Rathindranath Tagore on September 19, 1941, "I would only have accepted the responsibility if I could not have escaped it. At my time of life the natural desire is to lessen the burdens, not to add to them" (Gandhi 1998a: 357). For Gandhi, the poet was himself an "institution," and he even felt the loss of Gurudev in 1944 after three years of the death of poet (Gandhi 1999: 407–8).

Gandhi was fully aware of his and Gurudev's work in the place of world history, even before his launch of the Quit India movement in August 1942. He wrote on August 30, 1941, that "Sarvodaya is for the Udaya—rise—of Sarva—all. Gurudev also longed to serve the world through India and breathed his last while doing so. He is gone but his experiment is unfinished. His mortal remains are no more but his soul is immortal like *ours*" (Gandhi 1999: 276). However, after the death of Tagore on August 7, 1941, when he was invited by Shyama Prasad Mukherjee (president of All India Hindu Mahasabha) to preside over a meeting to pay tribute to Gurudev at Calcutta, he refused to participate because it conflicted with his engagement to raise money for Andrews's memorial fund (Gandhi 1999: 232).

Gandhi became a great world leader because of his experiments with truth through nonviolent political actions at South Africa (1893–1914), and in India through his programs of Satyagraha at Champaran (1917), Kaira (1918),

Ahmedabad (1918), and the noncooperation movement (1920–1922), the civil disobedience movement (1930–1934), and the Quit India movement (1942); ultimately securing freedom from the British Empire. Tagore was not alive to guide him in the Quit India movement, and Gandhi had distanced himself from him a few years prior when Subhas Chandra Bose contested the election against Pattabhi Sitaramayya at the Tripuri Congress session in 1939 and differed with him in noncooperation programs in South Africa, Champaran, Kaida, and Ahmedabad. He would have not become a great world leader if his programs had failed and not secured freedom from the British colonialism. It would be a great historical injustice to assess Gandhi as coming out of the womb of thought of Tagore and making Tagore essential to Gandhi in making him a world leader; which would also undermine the personality and words of Albert Einstein on Gandhi, that, "Generations to come, it may be, will scarce believe that such a one as this ever in flesh and blood walked upon this earth" (Einstein 1985: 78).

References and Further Reading

African National Congress (ANC). Mohandas K Ghandi. ANC website. Available at http://www.anc.org.za/ancdocs/history/people/gandhi/1-4.htm, accessed August 18, 2010.

Bandhu, Deep Chand, ed. *History of Indian National Congress, 1885-2002*. Delhi: Kalpaz Publications, 2003.

Bowles, Chester. *A View from New Delhi: Selected Speeches and Writings, 1963–1969*. New Delhi: Allied, 1969.

Brown, D. Mackenzie. *Indian Political Thought from Manu to Gandhi*. Berkeley: University of California Press, 1958.

Chandra, Bipan, Amales Tripathi, and Barun De. *Freedom Struggle*. New Delhi: National Book Trust of India, 2001.

Einstein, Albert. *Ideas and Opinions*. New York: Crown, 1985.

Gandhi, Mahatma. *An Autobiography or the Story of My Experiments with Truth*. Ahmadabad, India: Navajivan, 2001.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 7. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1995.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 8. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1995.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 12. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1995.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 14. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1998.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 15. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1998.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 18. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1994.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 20. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1995.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 67. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1994.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 74. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1998.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 76. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1998.

Gandhi, Mahatma. *The Collected Works of Mahatma Gandhi*, Vol. 77. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1999.

Gandhi, Mahatma. *The Gandhi Reader: A Sourcebook of His Life and Writings*, edited by Homer A. Jack. Bloomington: Indiana University Press, 1956.

Gandhi, Mahatma. *Satyagraha in South Africa*. Ahmadabad, India: Navajivan, 2003.

Gandhi, Mahatma. *The Selected Works of Mahatma Gandhi*. Ahmedabad, India: Navajivan Publishing House, 1968.

Gandhi, Mahatma. *Selections from Gandhi*. Ahmedabad, India: Navajivan Publishing House, 1948.

Prasad, Devi, and Rabindranath Tagore. *Philosophy of Education and Painting*. New Delhi: National Book Trust, 2001.

Raghuramaraju, A., ed. *Debating Gandhi*. New Delhi: Oxford University Press, 2006.

Rolland, Romain, and Mahatma Gandhi. *Romain Rolland and Gandhi Correspondence (Letters, Diary Extracts, Articles, Etc.)*. New Delhi: Publications Divisions, Ministry of Information and Broadcasting, 1990.

Suda, Jyoti Prasad. *Main Currents of Social and Political Thought in Modern India,* Vol. 1: The Liberal and National Tradition. Merrut, India: Jai Prakash Nath, 1963.

Tagore, Rabindranath. *The English Writings of Rabindranath Tagore: A Miscellany*. New Delhi: Sahitya Akademi, 1996.

8

Sacco and Vanzetti were innocent.

PRO Annessa Babic
CON Arthur K. Steinberg

PRO

After the 1917 Bolshevik Revolution, communism no longer lingered as a theory or ideological threat. The turnover of Russia's regime its withdrawal from World War I—which coincided with the U.S. entry into the European Front—led to pandemonium throughout much of the Allied countries. The United States experienced this fear and hysteria of communism, its perceived infiltration into U.S. society, and its effects on political and moral order. The United States had a small number of radicals; most notably a small number of groups that caused a series of bombings in 1919. In April 1919 the post office intercepted several mail bombs, and in June eight bombs exploded in eight cities within minutes of one another. This suggestion that there was a national conspiracy of revolution and communism led 30 states to enact peacetime sedition laws. Revolution, which had been the founding of the American republic, now became a crime on the local and national levels.

These acts of suppression continued to escalate as Attorney General A. Mitchell Palmer and his assistant J. Edgar Hoover led a series of raids on suspected radicals beginning New Year's 1920. Approximately 6,000 people were arrested, most released, but about 500 noncitizens found themselves deported. The fears of this scare quickly subsided, as these things often do, but in May 1920 some of the greatest effects of the Red Scare erupted in Braintree, Massachusetts. There, police hastily charged two Italian immigrants with the murder of a paymaster and his guard. This murder, and subsequent arrest, not only put Nicola Sacco and Bartolomeo Vanzetti in history books but it also shocked the nation and world.

Early Life

Sacco immigrated to the United States at 17 from Torremaggiore, Foggia, and Vanzetti came to the United States from Villafalletto, Cuneo. They came to the United States in 1908; they did not meet until 1917. Circa 1917 the two immigrated to Mexico to avoid conscription into the army during World War I, and they came back to the United States in 1920. The duo settled in Massachusetts

171

and quickly made roots in the local anarchist community. They most likely met through anarchist meetings under the discipline of Luigi Galleani. Galleani was an Italian anarchist who called for revolutionary violence. A handful of his supporters were arrested or questioned for their activities, and just two days before their arrest, Andrea Salsedo fell to his death while being questioned. This incident, along with the intense antianarchist campaigns spreading throughout the country, led many Galleanists to go underground for their own protection.

Politics of the Moment

As Sacco and Vanzetti were charged with the murder of the Braintree paymaster, the country stood in the midst of its own senses of social chaos. The ending of World War I brought more than just the fear of communism infiltrating and destroying American life. On January 16, 1920, the 18th Amendment was passed, making prohibition a national law, and on August 26 the 19th Amendment gave women the franchise. Also in 1920 the United States saw more urban dwellers than those in rural communities for the first time, and conservative Warren G. Harding won the presidency. Amid these political moments the gross national product stood at $91.5 billion, up from $35.3 billion in 1910, and the population rose to 105,710,620 (from 91,972,266 in 1910, a 14.9 percent increase). The increasing population in 1920 was about 5 percent more than the 1910 figures, but might have been greater had it not been for the immigration restrictions during the war years and restrictive immigration acts within the United States. Beginning in 1921 the U.S. Congress began systematically to restrict immigration from nonwhite, non-European countries. As early as 1790 Congress passed laws concerning newcomers, but the Immigration Act of 1921 restricted immigrants to 150,000 and 2 percent of the 1890 census. The 1890 census served as the starting point for this legislation because Frederick Jackson Turner declared the frontier closed as of 1890. Also, the 1890 census showed few, if any, immigrants from Asia and southern Europe. These pieces of legislation directly correlate to the rise of nativism inside the United States.

The ferocity of World War I birthed a new sense of patriotism throughout the United States. German Americans, socialists, anarchists, communists, and anyone who did not fit the ideal of the mainstream American faced the ugly wrath of suppression and becoming a target. The Industrial Workers of the World (IWW) and the Socialist Party were the center of many grassroots political discussions of the era. The IWW, primarily located in the northwest, fell under the leadership of Bill Haywood; Eugene Debs led the Socialist Party. These groups openly opposed U.S. entry into the war, and accordingly they became targets for local and federal campaigns seeking out individuals acting unpatriotic or un-American. In January 1919 building hostilities surmounted, and on January 21, 35,000 shipyard workers struck. In February a total of 60,000 Seattle area workers joined the ranks of

strikers, and the sheer number and quick rise of unity among the workers accelerated already rising tensions throughout the nation.

The Seattle strike represented what World War I brought to the warfront and what the postwar world meant to the worker and typical American. Life and war were now mechanized, death tolls were higher, and as in the case of the strike, workers joined in larger groups of solidarity to voice themselves. In the initial shock of the strike, most denoted it at as a Wobbly action, but in later years that was reputed. Seattle Mayor Ole Hansen announced that he had equal numbers of city policemen and federal troops to quell the labor unrest. This announcement came from the fear that Seattle would be paralyzed from the strike. These claims circulated throughout national media outlets, and pressure mounted in Seattle to end the strike. On February 10 the strike ended because its leaders saw the potential for it to damage the labor movement and be unsuccessful. The mayor took credit for the strike, and he later used it as part of his public speaking engagements about the importance of quelling communism.

A series of bombs continued to fuel the communist fires in the United States. On May Day 1919 rallies and riots occurred throughout the country, with Boston, New York, and Cleveland hit the hardest. In April and June bomb plots were thwarted, but media coverage of the events increased the perceived need to seek and stop enemies and anarchists. As the fall of 1919 approached, tensions did not ease the average American's mind. In September the Boston police force went on strike, which fueled a panic that the Boston police force was being taken over by the "Reds," which not only overtook the city but also the nation. Within four days, the police commissioner announced that the striking police would be returning to their jobs, the strike ended, but labor unrest did not. In September, steel workers united across the country. The strike began with 275,000 and quickly rose to 365,000 employees. Yet, even with these large numbers of rising strikers, steel owners held fast, and by January 1920 fewer than one-third of the strikers held fast to their positions. The strikes ended without the owners conceding to a single worker demand.

In the midst of these strikes U.S. veterans united under the new front of the American Legion. The legion, which is now the largest and most active veterans' organization, began in Saint Louis in 1919, under the auspices of protecting veterans, promoting the nation, and seeking to lead the front against the "Reds" in the United States. The legion quickly built itself up as presence throughout the nation, and by the mid-1920s, fighting Reds became synonymous with the legion and the federal government's new offices of the Federal Bureau of Investigation (FBI). The FBI compiled over 200,000 cards on radical organizations throughout the decade, and on January 2, 1920, the Palmer Raids arrested thousands in 33 cities.

By the spring of 1920 the national mood began to soften, and media attention began to shift its focus away from the suppression of Reds. In May several prominent attorneys issued their report about violations of civil liberties. Some

of these attorneys were Harvard professors Dean Pound, Zachariah Chaffee, and Felix Frankfurter. Frankfurter later became a famed Supreme Court justice. This report also arose from New York State, removing several elected socialists from office, and momentary voices of reason continued to be heard as the fear of communism bred the horror of censorship, alienating the larger U.S. public. Unfortunately, these voices of reason did not come soon or loud enough for Sacco and Vanzetti.

The Charges, Trial, and Conviction

On April 15, 1920, F. A. Parmenter and Alessandro Berardelli were gunned down in South Braintree, Massachusetts. Parmenter was a shoe factory paymaster, and Berardelli was his guard. Initial reports claimed there were two gunmen, who escaped capture, and reports noted that they took about $15,000 in cash. The crime itself was not that uncommon in postwar America. Robberies and other acts of lawlessness were common in working-class towns and communities in postwar America, as individuals scrambled to make ends meet, and criminal subcultures rose in the era.

Originally Sacco and Vanzetti had not been suspects in the crime, but three weeks after the shootings they fell into a police trap intended to nab a suspect in another Braintree crime. They were arrested at Boda's Garage, while attempting to obtain a car that fit the description of one seen near the crimes. The pair, both carrying weapons at the time of arrest, lied to police about their whereabouts and weapons. The basis of their arrests resided on the facts that Sacco owned a pistol like the one used in the Braintree shooting, and the two were attempting to reclaim an automobile that had been seen in the area of the murders. They knew little English, and they most likely lied to hide their Italian heritage. Racism raged in the United States, and prohibition had led to the quick rise of Italian mafia groups. Many Italians changed their names and continually fought the stereotype of being mobsters. They were officially indicted on September 11, 1920. Vanzetti was also charged with a December 24, 1919, robbery in Bridgewater. With the pending charges of the Braintree murders looming over his head, Vanzetti first stood trial for the Bridgewater crime in the summer of 1920. His standing trial for the crimes separately broke Massachusetts's court tradition, but as history would show, the Braintree trials were about to make national and international headlines. With historian's hindsight, the outcome of the Bridgewater trial set the tone and precedent for the Braintree crimes. Vanzetti was found guilty of the botched Bridgewater robbery, even though he had a sound alibi and witnesses to attest that he was not there. He received a harsh sentence of 10 to 15 years for attempted robbery with no injuries.

After the shockingly harsh ruling in the Bridgewater case, Fred H. Moore was brought in as the defense attorney. Moore suited the bill for a defense

Bartolomeo Vanzetti (center left) and Nicola Sacco (center right), immigrant Italian anarchists convicted in a 1920 armed robbery and murder, enter the Dedham, Massachusetts, courthouse in handcuffs. (Library of Congress)

attorney on the case as his work on the West Coast in labor disputes with the IWW and the Ettor-Giovannitti case brought him fame. The recently formed American Civil Liberties Union hired him for his radical nature and political expertise. The latter reference derived from a 1912 textile strike in Lawrence, Massachusetts. When Moore took the case, he pulled upon his own socialist beliefs and the rising fear of anarchists in the United States. He grounded his defense on acknowledging the anarchist activities of the Sacco and Vanzetti, arguing that their arrests derived from political motives over judicial ones. Moore believed that the prosecution was following suit of federal authorities and targeting radical groups and individuals who did not blend with the U.S. ideal. During the trial, he attempted to connect the actions of the Braintree police to the likeness of the Palmer Raids in 1919. Moore utilized familiar grassroots techniques to build a public defense for Sacco and Vanzetti. He organized meetings, began new investigations, and pulled for union and international support of the duo. He even enlisted the help of the Italian government, much to the chagrin of U.S.-based anarchists. Although Moore's efforts would be in vain, they did turn the case into an international cause.

The actual trial lasted six weeks (beginning on May 21, 1921), a lengthy time period indeed, and on July 14, 1921, the jury found Sacco and Vanzetti guilty.

Their guilty verdict came in the face of conflicting evidence; testaments to gang activity in Braintree (most notable the Morelli Gang), and a confession by Celestino Madeiros (a convicted bank robber) in 1925 to the Braintree murders did little for the defendants. Of the eyewitness accounts of the murders, the testimony fell apart on cross-examination. More so, the defense had several accounts stating that Sacco and Vanzetti were in Boston when the shootings occurred. To compound the controversy of the case, the prosecutor—Frederick G. Katzmann—continually brought the defendants' political actions into the courtroom. Clearly, he attempted to use politics as a motive for conviction.

The prosecution had seven eyewitnesses stating that Sacco was seen with the bandits after the Braintree murders, and when the defense cross-examined these individuals, none of them could concretely state they saw either defendant in the area of the crimes. As a side note, none of the witnesses were required to pick Sacco out of a lineup, and most could not identify him. Ballistic experts for the defense testified that the third bullet could not have been fired from Sacco's revolver, and the prosecution had a witness who attested that the bullet's scratches came from a defect in the Sacco's Colt chamber. Prosecutors claimed that a cap, with earflaps, belonged to Sacco. It was found at the scene, and testimony attributed its hole from the nail he hung it on at his place of employment. Sacco, in contrast, claimed that he did not own such a cap and when he tried it one he stated that it did not fit. The prosecutor argued that it did fit. Sacco testified that he was in Boston, at a consulate clerk's office, on the day of the shooting. The clerk, who saw several people per day, could not remember him. The prosecution hounded on this lack of memory as evidence that Sacco lied about his whereabouts, even though seven witnesses claimed seeing Sacco in a Boston eatery.

The evidence for and against Vanzetti followed the same trends at that for Sacco. Four witnesses placed Vanzetti at the crime scene; with one saying he was spotted at the train station, one saying he was driving the car, one declaring he was in the passenger seat, and one placing him in the back of a car a few blocks away. At the time of arrest Vanzetti carried a .38 caliber Harrington and Richardson revolver, which was also the weapon the guard had carried. Both guns had repaired hammers. Vanzetti claimed to have bought the piece at a local store, but he could not remember the name of the shop. Yet, during the trial Vanzetti testified that he bought the weapon from his friend Falzini for protection. He claimed the rise in hold-ups made him fearful of carrying cash from his fish-peddling business. Vanzetti testified that he went to Bridgewater to collect anarchist literature, but as with the gun story he could not remember where he went. At the time of his arrest he claimed he was in Bridgewater to collect the car that he and Sacco were arrested trying to retrieve. Moore omitted from trial testimony that Sacco and Vanzetti were actually retrieving the car so they could collect dynamite. The defense felt this revelation would have fueled the political fires already enveloping the case.

Even more controversial than the conflicting testimonies and lies of the defendants was the behavior of Judge Webster Thayer. He had also sentenced Vanzetti to the extremely harsh ruling in the Bridgewater case. Judge Thayer made his contempt for the Italian duo evident as he continually threw out evidence and testimony in the case. The statements he threw out were from known anarchists and immigrants, mostly Italian. He stated the Madeiros confession was not substantial evidence to acquit Sacco and Vanzetti, and he vehemently declared it did not justify a new trial. Judge Thayer's actions demonstrated the rise of nativism in the United States, and his clear disdain for the Italian men can be linked to the continual and rising racial relations circumventing the nation. The Massachusetts Supreme Court reviewed the case but ruled in favor of the prosecution in 1926 (*Commonwealth v. Sacco and Vanzetti*, 255 Mass. 369, 151 N.E. 839 [1926]). Attempts to save the duo lasted until 1927.

Throughout the trial, and after, the actions of Moore captivated the local and national media. His use of media foreshadows later 20th-century trials that utilized the media to pull on public opinion and racism. Most notably, a comparison can be made to the use of media in the 1995 murder trial of O. J. Simpson. Simpson was acquitted of the 1994 murder of his ex-wife and her friend, but the use of media outlets certainly compares the sensationalism of the Sacco and Vanzetti case. Also, Simpson was asked to try on a pair of gloves (found at the crime scene) and defense attorney Johnnie Cochran famously quipped "if they do not fit you must acquit." As the prosecution argued the cap fit Sacco, Simpson's prosecutors claimed the same.

Judge Webster Thayer

Webster Thayer was the judge who presided over the Sacco and Vanzetti case. Born on July 7, 1857, at Blackstone, Massachusetts, his father, Samuel Thayer, was a butcher. Educated at Worcester Academy and Dartmouth College, he then became involved in state politics, initially as a democrat and then as a republican. In 1917 he was appointed a judge of the Superior Court of Massachusetts.

Before his involvement in the Sacco and Vanzetti case, Thayer had been particularly critical of anarchists. At the end of a trial of Segris Zagroff in 1920, Thayer was angry that the jury had acquitted Zagroff, a self-confessed anarchist, and rebuked the foreman of the jury, claiming that the jury had clearly not listened to the evidence.

During the trial of Sacco and Vanzetti, Thayer was once again accused of bias, and his actions were widely criticized. However, unlike the Zagroff case, some of the jury went so far as to praise Thayer. After the execution of Sacco and Vanzetti, Thayer's house was bombed on September 27, 1932, and his wife and housekeeper were both injured. He moved to the University Club in Boston where he remained under guard until his death on April 18 of the following year.

Moore raised the large sums of money needed to utilize the media largely from working-class groups, and when he offered a reward for information as to the real criminals, the defendants even questioned his strategy. In 1924 Moore was ousted as defending council, to be replaced with William Thompson. Thompson, a Brahmin, was well respected in Boston and he sought to protect the reputation of Massachusetts law and the defendants. Initially, he did not care for the political actions and opinions of the duo, but as the trial progressed he publically stated that he deeply admired their dedication and determination. Thompson attempted to down-play the political natures of the case, but the fires on this front had already been ignited.

Political and social circles continued to analyze the case, and law professors continually used the case in lectures and briefs. Felix Frankfurter, future Supreme Court justice, saw the case a test of the U.S. justice system. As the Sacco-Vanzetti appeal waned, on protestors, conservative and liberal alike, stood in defense of the defendants. These individuals rallied under the guise of protecting the American ideal of innocent until proven guilty.

Even after this much publicized retrial Sacco and Vanzetti were still found guilty and finally sentenced to death on April 9, 1927. Public outcry—and international protests in cities like Paris, London, Mexico City, and Buenos Aires—caused the Massachusetts governor to temporarily consider a clemency for the two men. He went so far as to appoint a committee on the matter—dubbed the Lowell Committee—and the committee claimed that the trials and judicial process had been fair and just in regard to Sacco and Vanzetti. The Lowell Committee got its name from A. Lawrence Lowell, president of Harvard University, its most prominent member. Members of the committee remarked that witnesses seemed more certain of their testimony during the trial than they did immediately after the crimes. As criticism rose, Harvard alum even dubbed the university as "Hangman's House."

The duo faced their fate on April 23, 1927, when they died in the electric chair.

Aftermath

In the aftermath of the trial in Boston, the nation and the international community looked on with awe, shock, and wonder. After their execution, Sacco and Vanzetti received an elaborate funeral, somewhat similar to the spectacle seen with the initial martyrs of the Boston massacre. Sacco and Vanzetti's procession began on the north end of the city and ended at Forest Hill Cemetery. Their bodies were cremated there.

In 1927 the debate continued to boil about the innocence of the two men, vigilantism guised under the law and racist tensions continually colored the arguments. Yet, the horrors of the Sacco-Vanzetti case quickly faded into the

background of the U.S. psyche as labor disputes increased and the economy tumbled. The October 18, 1927, coal strike forced the nation to begin reexamining itself. Colorado coal miners, under the IWW, joined in unison and urged miners across the nation to strike for their own rights and in protest of the Sacco-Vanzetti execution. In total 117 mines closed during the strike, and the federal government had to intervene on behalf of the mine owners. The fear of socialism continued to surge in waves, but by October 1929, the nation had to focus on the failing economy. Socialism, corrupt businesses, unchecked capitalism, and the government took parts of the blame. Yet, the uncertainty and disdain left behind by the Sacco-Vanzetti sensational trial did not quell racism and the Red scares of the post–World War II era from rocking the nation again.

Conclusion

Scholars and amateurs have continually reexamined the Sacco-Vanzetti trial. The most damning evidence surfaced in 1961 and 1977 with the aid of forensic science. In 1961 ballistic testing confirmed that bullet number three did come from Sacco's weapon, and in 1977 reports clarified that Bereradilli's revolver was a .32 and not a .38. This new evidence sheds a stronger shadow of guilt on Sacco, while seemingly exonerating Vanzetti. Scholars generally agree that Vanzetti was innocent of the Braintree crime, and that Sacco should not have been convicted under the circumstantial evidence given.

Yet, when Sacco and Vanzetti were convicted they did not receive a fair and just trial. They stood trial in a crowded court room, listened to prosecution witnesses give continually changing and flimsy accounts of the crimes, and watched as the reporters and onlookers spun the story for the biggest audience. When looking at the sensational nature of the trial, the onlooker can easily see how witnesses may have testified so they would have a slice of the glamour. More so, the rise of mafia units in the wake of prohibition—particularly Italian ones—heightened tensions against immigrants. Also, Italy stood as one of the causes of outbreak of World War I, which increased shaky Italian race relations in the United States. It is clear that in 1920 and 1927 the evidence did not support a conviction of Sacco and Vanzetti. At the time, the duo merely stood as a national sacrifice to show strength and determination for quelling anarchists, communists, and activities not directly correlated with the American sense of democracy.

References and Further Reading

Avrich, Paul. *Sacco and Vanzetti: The Anarchist Background*. Princeton, NJ: Princeton University Press, 1991.

Bortman, Eli C. *Sacco and Vanzetti*. Beverly, MA: Commonwealth Editions, 2005.

Ehrmann, Herbert B. *The Cast that Will Not Die: Commonwealth vs. Sacco and Vanzetti*. Boston: Little, Brown, 1969.

Higham, John. *Strangers in the Land: Patterns of American Nativism*. New York: Atheneum, 1969.

Newby, Richard, ed. *Kill Now, Talk Forever: Debating Sacco and Vanzetti*, 7th ed. New York: Author House, 2007.

Sacco, Nicola, and Bartolomeo Vanzetti. *The Letters of Sacco and Vanzetti*. New York: Penguin Classics, 2007.

Selmi, Patrick. "Social Work and the Campaign to Save Sacco and Vanzetti." *Social Service Review* 75, 1 (March 2001): 115–34.

Topp, Michael M. *The Sacco and Vanzetti Case: A Brief History with* Documents. New York: Palgrave Macmillan, 2005.

Young, William, and David E. Kaiser. *Postmortem: New Evidence in the Case of Sacco and Vanzetti*. Boston: University of Massachusetts Press, 1985.

CON

The jury system entered American jurisprudence as a result of the Magna Carta of 1215, first signed by an English sovereign, John, and his assembled nobility. John had violated many of his common law obligations to the royalty of the English realm. The duties and obligations agreed upon by king and nobility under papal compulsion became a contractual agreement, and they are the foundation of the "due process" system memorialized in the American Bill of Rights and the U.S. Constitution.

An element of the Magna Carta was the granting of a criminal trial by jury of one's peers, usually composed of 12 men impaneled according to jurisdictional methods that do not violate the Constitution. The 12 chosen from a "pool" were originally of like kind in England and the American colonies; they were neighbors probably of similar trades and religion. Since their selection was from a small geographic area, their interests and prejudices were common. When an individual comes to court for a jury composed of his "peers," the law protects him.

Sacco and Vanzetti arrived in Boston, immigrants from Italy ca. 1919. They represented the new immigration, people from southeast and Eastern Europe whose whole culture was quite different from that of the old immigration. Their culture, habits, food, language, and smell caused consternation among the Brahmin culture of staid Boston. Massachusetts was not prepared for the self-proclaimed radicals espousing violence and revolution, especially from two Catholic Italian immigrants whose manner and presence were repugnant to the Boston Protestants.

Local society easily blamed these self-proclaimed disseminators of violence as tools of the same movement caused by violent immigrants from the nascent Soviet Union's godless communism. Americans remembered the Great War and

tsarist Russia's treachery in unilaterally withdrawing from the war to the detriment of the Christian West. The United States had been so concerned that she sent aid to several monarchist tsarist Russians in an effort to defeat Lenin and his cohorts in their civil war. It was these anarchist immigrants who believed in terror and violence as a way to replace Western values with communism and bolshevism.

A. Mitchell Palmer, President Wilson's attorney general, used his federal legal authority to wage a war against the disturbers of American society. He feared the fabric of society was endangered. With wide public support, he attacked venues he perceived as propagating alien values. He used extrajudicial means and detained people without warrants, smashing offices of groups he believed subversive without any semblance of due process. In 1919 he instituted the Palmer Raids and arrested 249 alleged revolutionaries and anarchists who were placed on the "Ark," the *Buford*, against their will, for immediate deportation.

Societal fears were not confined to any one group. A man was sentenced to six months in jail without due process for stating Lenin was not stupid. In Chicago a sailor shot another who did not rise to the national anthem during this period, the Red scare. Anti-immigration sentiment was common. Sacco and Vanzetti were two Italian laborers who had immigrated to the United States before the political and social atmosphere became anti-East and-southern European.

The Sacco-Vanzetti affair erupted at a time when many "real" Americans opposed the influx of strangers. A murder occurred in Boston on April 15, 1920, where two unidentified men in Braintree, a Boston suburb, committed a robbery and murdered one security employee and wounded another responsible for the Slater and Morrill Shoe Companies payroll. The evidence indicated that on Thursday at 9:18 A.M. a train arrived carrying a strong box containing a $30,000 payroll for the plant employees. Shelly Neal, an American Express employee, transported the box to his office on Railroad Avenue. He observed a touring car; he thought it was a Buick, and was not certain as to its make, year, or license number. He never informed the court or attorneys to whom he made the observation because he should have been concerned with delivering the box contents safely and couldn't guarantee the veracity of his observations. He delivered the box to Frederick Parameter whose office was in the company building. The plan was to deliver one of the canvas bags to the shoe company management building.

Neal took the bag for the manufacturing plant up an internal staircase and left it with Parameter. Prior to delivery to the factory, he allegedly divided the $15,776 among the 700 factory workers, but no testimony was introduced that the funds had been distributed to the various employees. Could Neal and his possible confederates have taken the funds or part of the amount for their private needs? At 3:00 P.M. Parameter and guard Alessandra Bernardelli took their usual route up Pearl Street with the factory funds. Parameter carried a .32 calibers Colt and Bernardelli a .38 Harrington.

They had to cross railroad tracks, and as they waited for the guardrail to rise, two men appeared with drawn revolvers. Parameter died and Bernardilli received four bullet wounds. The two unidentified men seized the two payroll boxes and climbed into a waiting vehicle that already had two passengers. As the vehicle sped away, someone fired shots in the direction of the getaway car.

None of the security people had their employment records examined for either a criminal or fiscal background check. Evidence may have indicated financial problems existed for one of the parties involved, an error committed by the defense attorney. Witnesses told police that the robbers and killers spoke Italian. None of the witnesses indicated they knew Italian or any language used in southern or East Europe, but did allege that the two fled in a Buick. Was it blue? What model was it? An Overland? James Bostock gave the police four shell casings with the hope that ballistics would identify the murder weapon, but no indication was presented that it belonged to Sacco, Vanzetti, or who fired it. No evidence was offered by the defense alleging that someone had stolen the guns.

Two days after the robbery and murder, two hunters discovered an abandoned Buick in the woods without license plates with the rear window shattered. No evidence was offered as to the reasons for the window's condition; there were no shell casings in or around the vehicle but a spare set of tires lay next to the car. Someone called the police to report a stolen and damaged vehicle. Police Chief Stewart believed the Braintree robbery and murder represented a pattern; a failed burglary occurred with a blue car on December 24, 1919. He sought to link the two crimes, but his opinion was conjecture. Again, unidentified witnesses swore the robbers spoke Italian. Their general description was "similar" to the appearance of Sacco and Vanzetti.

The chief, aware of the impending deportation of Ferruicco Coacci who also worked at Slater and Morrill, knew this admitted anarchist was scheduled to leave the United States on April 15. Coacci waited until April 18. Stewart believed he was involved in the Braintree incident because he was eager to return to Italy. The chief surmised Coacci participated because he had no money, which would provide the motive for his actions. There was also no evidence linking him to the Buick nor was any money found on him.

Stewart went to Coacci's rented house without a search warrant and met Mario Buda who informed the police of a Buick in a local garage. It was not being repaired. Stewart left and returned on May 1, but found everyone gone.

Chief Stewart persisted and located the car, an Overland, at the Elm Street garage where he found a Buick with an undamaged rear window. He asked the owner's wife to call the police when someone came for the vehicle. She did call Stewart secretly, she thought. On May 5, Buda appeared with three Italian friends, none of whom were described as Sacco and Vanzetti. They became nervous. There is some question whether Buda knew of the call to the police. Buda and another person hastily left the garage on a motorcycle. The others,

Sacco and Vanzetti, took a streetcar but were arrested at the station because of Stewart's fast work.

They were taken to the police station and subjected to a lineup but were not identified as having been involved in Braintree. They spoke English, but to what degree is not known. When arrested, they carried anarchist materials and Sacco had a .32 Colt revolver. Even though Vanzetti had a weapon he denied, the police found it. His reason for carrying the weapon was that he was a fish merchant who carried large amounts of cash. He feared being robbed. When asked why he lied, he did not have an answer. Stewart now believed he had enough evidence to connect Sacco, Vanzetti, Buda, and Coacci with Braintree. Stewart then obtained a warrant.

Obtaining a warrant was only the beginning of building a case; Stewart realized he needed someone more skillful at interrogation. The state prosecutor, District Attorney Frederick Katzmann, conducted the interrogation. Katzmann had no evidence that Sacco had fired his .32 Colt, and Vanzetti again informed the prosecutor he had a gun to protect himself because he was fish peddler. Sacco and Vanzetti denied knowing Buda and Coacci or being at the Elm Street Garage. Sacco stated that he had been at work the day of the robbery, and Vanzetti alleged he could not remember what he had done that day or where he had been. This evidence was used against them in prosecuting the case.

On September 28, 1920, Sacco and Vanzetti were arraigned in the Dedham Courthouse and both pled not guilty. Bail was not requested and probably would have been denied because of the capital crime involved and their reputations. As a result, they remained in jail until some future date. While in jail, Moore, their attorney, spoke with Sacco's cellmate and obtained an affidavit from several inmates who said Sacco was not guilty. Sacco's cellmate Celestino F. Madeiros provided him with a note stating Madeiros had been at Braintree during the robbery and that Sacco was not there. If true, why would Madeiros involve himself since he certainly knew the penalty for murder would be execution if found guilty? Advocates of Sacco and Vanzetti's innocence told Moore the Braintree robbery was similar to those committed by the Morelli gang, but no investigation was conducted. An unidentified party informed Moore that he was in the Buick and Sacco and Vanzetti had nothing to do with the robbery and murder.

Judge Webster Thayer requested the assignment of presiding over the Sacco-Vanzetti trial from the senior circuit judge in his district. He had graduated from Worcester and Dartmouth College and represented the Boston Brahmin class. He assumed the bench in this case in 1920 and once on the bench continually violated judicial canons pertaining to judges by his public expressions and attitude in the case: the fear of anarchism, destruction of the culture, socialism, and immigration, all of which he believed were threatening. His bona fides included admiration for Attorney General Palmer, who conducted raids to purge America of dangerous alien influences. On numerous

occasions, in and out of court, Thayer expressed less than objective views about the innocence of Sacco and Vanzetti. He cautioned that the defendants were a threat to national security. As the trial proceeded, Thayer was not shy about expressing his attitudes publicly about the defendants, saying they should get what they deserve. However, attitudes toward his conduct were not universal. The *Boston Globe* supported the judge's conduct.

As trial judge, Thayer was responsible for the conduct of the trial and his expressions. He controlled the admission of evidence such as not to introduce facts detrimental to the character of the defendants. That they were active anarchists did not pertain to their trial for murder. He heard motions and either granted or denied them when legally relevant to the case (such as their having fled to Mexico to avoid the draft), controlled the conduct of the jury, attorneys, and the courtroom, and not express his personal views during the course of the trial. His denunciations included antiliberal comments, constant admonishments of the defendants' attorney, and denouncing a jury for acquitting Sergei Zagroff, an anarchist in April 1920. The judge had also tried Vanzetti for robbery in December 1919. This was extremely onerous because the defense attorney at the probable cause hearing had not offered any evidence to defend his client.

Fred H. Moore came from the West Coast, which had a culture quite different from Boston's. His reputation was enhanced by his appearance; he walked the streets of Boston in a broad brimmed cowboy hat and sandals. His reputation was of an excellent and passionate orator and defender. His conduct offended many staid Bostonians, and his outbursts in court were not uncommon. They proved effective for a time. He knew how to tailor comments to particular jurors and understood their personalities. His personality and Thayer's clashed constantly; they were constantly attacking each other. Motions, rulings, and comments about the admissibility of evidence and witnesses by the judge to the jury were tainted in many instances. Moore's objections were frequently overruled. Moore had represented Industrial Workers of the World (detractors dubbed the group "I Wont Work") members, who were also socialists and willing to use violence to make their points. This presence further irritated the judge who abhorred them.

Fred Moore met Sacco in jail on several occasions. Moore's career of defending socialists provided confidence. His approach for defending his client was not particularly orthodox for Boston. He intended to use public relations and he had the Sacco family available at all times for sympathy. Moore helped establish a defense fund, but the inability of the defendants to speak English at a truly coherent level made communications and understanding a problem. He sought the aid of organized labor and anarchist movements. Moore contacted Harvard Law professor Felix Frankfurter and Roger Baldwin, the head of the newly formed American Civil Liberty Union. Moore's tactics relied heavily on public support for his clients. He presented himself at every opportunity to any group that would listen to his case. In his efforts he found two people

---

Felix Frankfurter and the Defense of Sacco and Vanzetti

Of the many people who came to the defense of Sacco and Vanzetti, one of the most famous was Felix Frankfurter. Frankfurter was born in 1882 in Vienna, Austria, from a devout Jewish family, and when he was 12 his family moved to the United States, settling in New York. He graduated from the City College of New York and went into law, becoming an assistant of Henry Stimson, the U.S. attorney for the southern district of New York. Stimson was, soon afterward, appointed as secretary of war by Howard Taft. Frankfurter's views were more radical than those of Stimson, so he was forced to keep his own counsel.

After World War I, Frankfurter came to be interested in trade unionism and in radical politics and had sympathy with some labor leaders. In 1920 he was a cofounder of the American Civil Liberties Union and urged for a retrial of both Sacco and Vanzetti, being sharply critical of the role taken by Webster Thayer, the judge in the trial.

Felix Frankfurter became a close adviser of Franklin Delano Roosevelt after the 1932 presidential elections and in 1939 was appointed, by Roosevelt, to the Supreme Court. He remained an associate judge of the Supreme Court until 1962, being a constant supporter of judicial restraint. He died in 1965.

---

who stated that Vanzetti didn't look like the killer, but they refused to be subpoenaed.

Sacco and Vanzetti did not reject a jury trial. While its selection was unique, no evidence was introduced that any of the jurors had preconceived notions about the innocence or guilt of the defendants. Judge Thayer, in his haste and showing a lack of patience, literally pulled potential jurors off of the street. All were subjected to *voir dire*. The 12 reflected attitudes in the region. There were no Italian or union members. Fred Moore, the defendants' defense attorney, drew men from various jobs, hoping to draw a sympathetic jury. Given the political atmosphere, it was impossible. By June 4,500 men had been *voir dired* and only seven accepted. The sheriff was ordered to gather another 175 prospective jurors, bring them to court, and finally five were selected. As the trial proceeded, the jurors were not allowed any accommodations. Since they were sequestered, they didn't have clean clothes or toiletries, and deliberated in uncomfortable surroundings.

Fredrick Katzmann again appeared, as state prosecutor, opposing a change of venue, and was favorably disposed when he heard that Judge Thayer would preside. As a result of an interview with the defendants in jail, Katzmann had obtained information he could use at the trial. The state sustained the burden of convincing the jury "beyond a reasonable doubt," proving the defendants were the perpetrators of the crime for which they were being charged.

Prior to trial, Katzmann and Stewart rechecked the guns, bullets, shells, and cap found as evidence in the crime. They interviewed over 200 witnesses.

Issues dealing with language were added to the mix. Most interviews concluded that Sacco and Vanzetti were not the murderers. Concerning identification, Sacco didn't smoke and another stated one of the murderers had a mustache. A witness informed the interrogators that Sacco couldn't stand blood and others said he was not the killer type.

On September 11, 1920, Sacco and Vanzetti were indicted for the robberies and murder at Braintree. Five days later, Buda, a suspect, went to Manhattan and procured a wagon and a horse. He packed the wagon with dynamite and 500 pounds of sash weights, wired a clock for noon, and departed. He left the cart in downtown Manhattan. When the bomb went off, cars were tossed through the air, trolleys blown off their tracks. Shrapnel from the explosion slashed those who were nearby. The area looked as if an unknown enemy had attacked it. Thirty-eight people were dead and dying while over 300 were taken to the hospital. The morgues were crowded. The authorities immediately blamed the anarchists, communists, and bolsheviki for the violence. To make matters worse, anarchist leaflets were found throughout the area.

Given the tense trial situation and the possibility of execution, Sacco and Vanzetti's memories and responses in court were confusing and didn't convince the jury of their innocence. Katzmann understood how to maximize human frailty and used it against the defendants. When they made contradictory statements, Katzmann decided it was due to their "conscience of guilt," a term never explained to the court. Katzmann's witnesses placed Vanzetti near the crime scene, but no evidence was introduced placing Sacco there. Mary Spaine and Lola Andrews made identifications, but not as to individuals. They would go no further than refer to a resemblance. They were certain that they had not seen the defendants shoot anyone.

The caliber of the weapon killing the guards was examined, but at that time ballistics science had not yet been perfected. Ballistics expert William Proctor could only say under oath that the bullet had been fired from a gun of that caliber. Katzmann did his best to confuse the jury as to minimize the testimony concerning the bullet's effect on jury deliberation.

On October 1, 1924, Judge Thayer heard a series of motions dealing with an abundance of matters. The court first ruled that the allegation that jury foreman Ripley had taken a gun into the jury room was hearsay. The court refused to enter this affidavit into evidence. At the same time, the other 11 jurors prepared affidavits, but Judge Thayer rejected the matter, stating "Damn them, they ought to hang them anyway."

Ray Gould, another witness, added little to the case. Moore referred to Roy Gould's indictment for a crime in 1911 for which he was never sentenced or served time. Moore discredited him, but he never served time. Moore successfully discredited other state's witnesses such as Melvin Corl.

Lola Andrews, a state's witness, identified Sacco and didn't deviate from her testimony, despite Moore's aggressive cross-examination. The examination

was heated and he attacked questionable evidence. She charged that Moore attempted to bribe her. When she said she was not sure about her identification, she alleged that the sheriff compelled her to identify Sacco and Vanzetti. Evidence was presented that she was being abused at work because she was accused of trying to help Sacco and Vanzetti, Italians. Moore then challenged her affidavit and the court did nothing. He then contacted the district attorney, who did not act. This was especially important since she had initially stated she looked out the window and witnessed everything. But no evidence was offered as to the time because the sun could have put a glare in the window and it may not have been possible to make a good identification because of it.

Thayer determined that the conflicting testimony about the Buick was due to a "consciousness of guilt" rather than evidence in the affidavits offered by witnesses. He rejected the affidavits, stating the jury had the right to believe the investigators.

The judge then denied Carlos Goodridge's motion alleging perjury, but provided no legal rational. He stated the affiant's character was in question since force was alleged to obtain the affidavit.

Louis Pelser's affidavit stated the defendant Sacco resembled the murderer, but couldn't say he saw him commit the murder. He then stated that he was forced under threat of perjury by the district attorney to identify Sacco and Vanzetti as the criminal perpetrators. Pelser retracted the affidavit he had provided Moore. This was after stating he saw nothing and after the jury had been taken to the crime scene where they were shown the bloody clothes of the two dead men.

Moore's cross-examination discredited several other witnesses' testimony. He debunked the idea that what may have been the murder weapon was stolen. He informed the court that the weapon had belonged to someone other than the defendants. A chain of custody for the evidence was not established, putting the existence of the real weapon in doubt.

All the affidavits garnered by Moore and Thompson were retracted. Albert Hamilton, a ballistics expert, said that Sacco and Vanzetti had been

Judge Webster Thayer (center) leaves the courthouse after denying the appeal of Nicola Sacco and Bartolomeo Vanzetti. (Library of Congress)

framed. Hamilton informed the court that Vanzetti's revolver did not have a new hammer. He used his scientific skills to show the court that ballistically it was not possible to charge Sacco with the crime. Hamilton even brought another expert to court from the Massachusetts Institute of Technology, whose own measurements indicated that the "so called mortal bullet never passed through the Sacco revolver." Hamilton's affidavit stated that an unknown person had tampered with his gun barrel. His affidavit, in conjunction with Proctor, was denied because it challenged government experts who worked for the U.S. Ordnance Department during the war. The judge informed the court that if the government's witnesses were not believable, then they would be guilty of perjury or malice. He went further by informing all parties that only by clear, satisfactory, and convincing evidence should a public official be challenged.

In his summation Katzmann worked with circumstantial evidence: Sacco was one of the shooters, Vanzetti sat in the Buick and was a conspirator; the prosecutor spoke about the defendants' "consciousness of guilt," their possession of anarchist material, flight to Mexico to avoid the draft, and that Sacco left his wife and child to fend for themselves. Katzmann's efforts were successful. The jury, no matter its origins, found the defendants guilty as charged by the indictments. The court then poled each juror, and each answered in the affirmative as to the question of guilt.

The defendants' only recourse was an appeal. Did they have the necessary grounds to appeal to the Massachusetts Supreme Court? Unfortunately for them, Judge Thayer would be instrumental in their efforts. For the next six years Sacco and Vanzetti's attorneys filed motions arguing that the witnesses were not credible or they were coerced; that there was evidence of jury misconduct, especially with the foreman and jury selection; there was a denial of due process; there was a failure of the court to allow the admission of testimony not known or accessible at the time of trial; among other salient evidence.

Judge Thayer was the court of first impression for the evaluation of the merits of appeal; since no judge wants to be reversed, how could he grant a new trial? The defendants filed their first motion in November 1921, arguing that the trial evidence was insufficient to warrant a guilty verdict. Thayer denied the motion, arguing that there was sufficient evidence and the jury had acted properly. There was no evidence about the *voir dire* of the 167 potential jurors other than the method of selection.

When the new evidence became available, Walter Ripley's conduct as foreman came into question. The defense attorney argued that Ripley had a weapon in his possession during the trial. Moore questioned the possibility that the jury during sequestration may have seen the .32 caliber weapon and shells. Viewing it and any discussion concerning it could have prejudiced them. Ripley also expressed a prior attitude about the guilt of the defendants. He told the other jurors that they were guilty. Since the court is bound to dismiss any potential juror

with preconceived views about the case prior to trial, Thayer was obligated to remove Ripley from the jury pool if the court knew. If Ripley had lied under oath, he was then guilty of perjury and should be punished. But Ripley died after the trial and there was no way of proving the assertions made by Moore.

Another motion supported by an affidavit from William Proctor, a ballistics expert for the prosecution, was presented to the court. Sacco's attorney argued that Proctor had misled the court. The court denied this motion when Proctor said the gun belonged to Sacco but there was no proof he fired it. There was no evidence introduced that the weapon had been stolen or lost.

The denial of the numerous Moore motions ended his participation in the case. The new attorney, William Goodrich Thompson, took the case believing that the defendants were being tried on circumstantial evidence. He recognized that he would have to deal with Thayer's prejudice; he believed there was no evidence supporting the accusations. His belief was buttressed by Judge Thayer's ranting about the case with Attorney General Crocker about an attitude that did not become public until much later. Thayer said, "They are not getting a fair trial but I am working it so that their counsel will think they are" (Watson 2007: 257). At a Dartmouth game, Thayer is to have said, "Did you see what I did to those anarchist bastards the other day? I guess that will hold them for a while! Let them go to the Supreme Court now and see what they can get out of them" (Avrich 1991:).

Moore and Thompson cited 34 errors in the trial. They indicated and questioned: Judge Thayer's trial conduct, conflict over ownership of the revolver/revolvers in question, the lack of identification of defendants, the testimony about their political activities, testimony about the ballistics identification of the bullet, and Judge Thayer's bringing their political activities to the attention of the jury. Thompson even compared the case to the infamous Dreyfus trial in France where an innocent man was convicted on circumstantial evidence.

The appeal was denied and the Supreme Court decision stated that all the cited errors were within the discretion of the presiding judge. This was despite the record, motions, and affidavits obtained by Moore and Thompson, Joe Morelli's confession, and officer Jacob's testimony that he saw Morelli in a new Buick in the neighborhood at the time of the robbery despite the controversy about the lack of a license plate and other identification.

After the appeal to the State Supreme Court was denied, the Massachusetts governor formed a committee to investigate the Sacco-Vanzetti case with Judge Thayer being the first witness. The governor understood public attitudes and prejudices against foreigners and his respect for the judiciary caused him to act gingerly. No evidence was introduced about the missing Buick or its connection to the defendants, Celestino Medeiros (who had confessed to the murders in 1925) died, and nothing remained as to his potential testimony, Thayer's rejection of Moore's and Thompson's motions were ruled within Judge Thayer's discretion, the ballistic conflict testimony was stayed, all despite the fact that

Massachusetts law stated that once a sentence had been announced there could be no stay of execution.

A total of eight motions had been filed concerning contradictory evidence. Any motions dealing with newly obtained evidence had to be submitted to the court for proof that it could not have been found at the time of trial. Attorney's incompetence or lack of due diligence would not be enough to reopen the case. In 1961 a ballistics test was conducted with new equipment indicating that Sacco's pistol had been used to murder the guard. Some authorities now believe that Sacco was guilty of killing the guard at Braintree.

They were executed on August 23. Two years after the execution Thompson interviewed Proctor and made inquiries about his responses in court. Proctor told him that Katzmann had prodded him when he refused to identify Sacco's gun as the one used in the murders. Proctor said in his interrogation that the murder bullet had come from a Colt, but he couldn't say with certainly that the bullet came from Sacco's weapon. Proctor signed an affidavit indicating that Katzmann had attempted to put words in his mouth. Proctor stated that if Katzmann had asked if bullet three had passed through Sacco's gun, he would have had to say no. The prosecutor's response was that the defense could have asked the question. Thayer had denied a motion to strike Proctor's earlier testimony, alleging Katzmann had conducted himself within the canons of the law.

The Sacco-Vanzetti trial represented the success of the U.S. juridical process. Despite Moore and Thayer's constant referral to their character, not admissible under all circumstances, the jury evaluated the evidence and came up with the correct verdict. They weighed the information provided by the state and defense and ruled.

Since the anarchists and pro-anarchist public fought for their release, they attempted to change the question of robbery and murder. Public opinion was the tool used to garner support for their cause. Few freedom advocates delved into the facts but instead relied on emotionalism in a time when bolshevism became a threat to the American way of life.

Their attorneys were not able to explain how the defendants knew the routines of the victims. Sacco was an employee of the Slater Company and received checks at designated a time. Who knew the path the guards would take to deliver the payroll better than an employee?

Stewart's instinct about Buda, the visit to his house, and Buda's reference to a Buick in a garage finally led Stewart to Elm Street where four men fled as the police arrived. Sacco and Vanzetti fled and were apprehended at a public transportation station, while their two colleagues fled on a motorbike, accounting for the four suspects.

Their interrogation by Stewart and Katzmann indicated Vanzetti's fears of having violating the law. The denied possession of a gun at a crucial time provided what the state needed to move forward. Combined with the ballistics

problems, the lack of knowledge and Porter's repudiation of his testimony raised doubts about his testimony.

Witness identification was a problem. Many of those questioned, on the stand and off, were not absolute in connecting the defendants to the crime. For example, identification from the window of the factory was not absolutely positive because of the sun and glare from the sun. But worse, Moore's affidavits revealed something wrong. Affidavits were recanted and others obtained allegedly under coercion. Moore, as a knowledgeable attorney, should have had a witness in the room when the affidavits were prepared to deal with that potential problem.

Judge Thayer's conduct inside and outside of court, while reprehensible, was not beyond legal canons. His comments at the club, statements in court, and admitted prejudice were beyond the pale of the law. However, the State Supreme Court's investigation found his conduct within the rules of procedure. Those attacking his conduct must remember they functioned a century ago and may have been the standard.

The ballistics finding in 1961 proved Sacco's revolver had been used to commit the crime. The results of the ballistics test indicate the State of Massachusetts was correct in executing Sacco and Vanzetti.

References and Further Reading

Avrich, Paul. *Anarchist Voices: An Oral History of Anarchism in America*. Oakland, CA: AK Press, 2005.

Avrich, Paul. *Sacco and Vanzetti: The Anarchist Background*. Princeton, NJ: Princeton University Press, 1991.

Frankfurter, Felix. *The Case of Sacco and Vanzetti: A Critical Analysis for Lawyers and Laymen*. Buffalo: Hein, 1927.

Goldberg, Jonah. "The Clay Feet of Liberal Saints: What George Clooney and Upton Sinclair Have in Common." *National Review* (January 6, 2006). Available at http://old.nationalreview.com/goldberg/goldberg200601061019.asp, accessed August 18, 2010.

Neville, John F. *Twentieth-Century Cause Celebre: Sacco, Vanzetti, and the Press, 1920–1927*. Westport, CT: Greenwood, 2004.

Sacco, Nicola, and Bartolomeo Vanzetti. *The Letters of Sacco and Vanzetti*. New York: Penguin Classics, 2007.

Watson, Bruce. *Sacco & Vanzetti: The Men, the Murders, and the Judgment of Mankind*. New York: Viking, 2007.

9

Warren Harding was murdered, rather than dying of food poisoning.

PRO   Elizabeth D. Schafer
CON   Kimberly K. Porter

PRO

President Warren Gamaliel Harding's sudden death on August 2, 1923, in a San Francisco, California, hotel stunned Americans. The next day, newspapers nationwide reported Harding's demise in front-page articles framed with thick black lines. As a train transported Harding's body from San Francisco to Washington, D.C., the press chronicled that trip then described associated memorial and funeral services that occurred during the following week. Photographs showed readers Harding's casket, First Lady Florence Kling Harding, administration officials, mourners, and other relevant images. Journalists had begun covering Harding's medical condition since late July when he had supposedly suffered food poisoning while on an official trip, according to his personal physician Charles E. Sawyer, who had given the press bulletins regarding Harding's health. After Harding died, newspapers printed Sawyer's explanation that Harding died from a cerebral hemorrhage without questioning him or the other physicians who had attended Harding in San Francisco about that diagnosis.

Accounts in the press at that time did not contemplate if any other causes might have been responsible for Harding's death. Initially, the grief-stricken populace expressed respect for Harding. However, speculation about why an autopsy had not been performed on the president soon provoked rumors that Harding, whose administration was characterized by scandals, might have been assassinated. People endorsing the belief Harding had been murdered identified various suspects, including Harding's wife, who had frequent access to the president and was often alone with him. Conspiracy theories also identified political enemies or individuals with motives to poison Harding. From the 1920s through the early 21st century, historians, journalists, and physicians, among others, have offered their opinions, examining feasible evidence as it became available. Some people have used this information to support their belief the 57-year-old Harding was killed, either through criminal malevolence or negligence.

193

Medical History

Medicine shaped Harding's early years, especially alternative medical homeopathic practices using natural substances in an effort to boost patients' immune systems. His father George Tryon Harding had attended a Cleveland, Ohio, homeopathic school and worked as a physician in their community, Marion, Ohio, where laws permitted individuals to practice medicine after completing registration in their local probate court.

In the mid-1880s, Harding purchased a hometown newspaper, the *Marion Star*, serving as its publisher and editor. Charismatic, Harding easily attracted friends and supporters who encouraged his civic interests and participation in Republican Party activities. Despite public approval, in his 20s the emotionally fragile Harding spent time in a sanitarium for nervous breakdowns. Harding began courting Florence Kling who taught his sister Charity piano lessons. The pair wed in 1891.

In 1896, Harding's mother, Phoebe Harding, acquired a license to practice medicine and tended patients with homeopathic methods. After a toddler died in Phoebe Harding's care in 1897, local physician Charles E. Sawyer, an 1881 Homeopathic Hospital College graduate, defended her medical capabilities and said she was innocent of harming the child, despite a pharmacist's analysis revealing morphine in the powdered medicine she had given the boy. Grateful for his intervention, the Harding family established a friendship with Sawyer.

Florence Harding enhanced her husband's publishing endeavors with her business skills, and the pair prospered financially. She encouraged him to seek political offices. In 1899 Harding was elected to the Ohio legislature. By 1903, Harding became the state's lieutenant governor. His wife suffered a kidney ailment, resulting in surgical removal of one kidney in 1905. Sawyer served as her physician and actively supported the Hardings' political goals. Harding won a seat in the U.S. Senate in 1914. Six years later, Harding secured the U.S. presidency. Florence Harding convinced her husband to appoint Sawyer as his presidential physician so that doctor could continue treating her. Adjusting to Washington, D.C., the First Lady consulted a local psychic, Madame Marcia, who predicted that Warren Harding would die abruptly while president. She suggested circumstances of his death would be strange.

In the fall of 1922, Harding named Navy Lieutenant-Commander Joel T. Boone, a Hahnemann Medical College graduate, as Sawyer's assistant. Boone had also completed graduate work at the U.S. Navy's medical school. Harding indulged in late-night drinking, tobacco use, and poker games at the White House and did not pursue exercise beyond rounds of golf. Stresses impacting the Harding administration, including labor strikes, aggravated Harding's increasingly fragile health. He continued to endorse homeopathic practices, writing a foreword for *American Homeopathy in the World War* (1923), edited by Frederick M. Dearborn.

Voyage of Understanding

Harding suffered influenza in January 1923, experiencing gastrointestinal distress, but resumed working before he recovered sufficiently, thus weakening himself physically. His wife had also been sick with a urinary tract infection and nephritis, swelling her kidneys. They both had access to homeopathic medications to mitigate pain associated with their ailments. Their illnesses concerned friends and associates especially when Harding decided to be the first U.S. president to travel to Alaska while in office. He envisioned his transcontinental trip, which he called a Voyage of Understanding, would enable him to interact with Americans and discuss his goals for the United States.

In the presidential railcar *Superb*, Harding and an entourage of cabinet members, staff, and assistants departed Washington, D.C., on June 20, 1923. He soon was immersed in a hectic schedule. Along the journey west, Harding presented speeches to crowds, sometimes in extreme heat or in downpours. Aware of two aides' suicides prior to the trip, Harding received more negative news during his travels of corruption by friends whom he had trusted. He was distressed that close associates serving in political positions he had attained for them had been disloyal to him. Harding suffered fatigue in Alaska, and the constant demands and shocks hindered his attempts to recuperate. He played cards, drank whiskey, and did not sleep much. Florence Harding briefly was sick in Juneau. Her previous nephritis attacks had worried Boone who had stored a coffin for her if needed on the ship USS *Henderson* transporting the presidential group between the mainland and Alaska.

Conspiracy Theories about President Harding's Death

It was Gaston B. Means in his *The Strange Death of President Harding* (1930) that raised the idea that President Warren Harding had been murdered. Means himself had been born in 1879 in North Carolina, the son of a lawyer. He graduated from the University of North Carolina and worked as a schoolteacher and then as a salesman before finding work with a New York detective agency.

With the outbreak of World War I, Means was hired to uncover plots about German involvement in espionage work. Soon afterward he was accused of enriching himself at the expense of a wealthy widow who died soon afterward. Means was cleared of her death. This did not stop him from being hired by the Federal Bureau of Investigation and then working bootlegging alcohol. His book on the death of Harding was a bestseller, but there were claims that it had been ghostwritten by a journalist named Mae Dixon. Means later became involved in the Lindbergh baby kidnapping of 1932, offering to act as a go-between with the kidnappers for the $100,000 ransom. He was later charged with grand larceny and sent to jail at Fort Leavenworth where he died in 1938.

While traveling on the *Henderson* to Vancouver, British Columbia, Harding consumed crabs, a gift from the Sitka, Alaska, community. On land, he golfed on a local course but soon felt nauseated and his chest hurt. Sawyer examined Harding, treating the president's increased breathing and pulse rates with digitalis, a medicine made from foxglove. Harding had suffered a finger infection on the ship and asked Boone to lance the wound, saying Sawyer's vision was not sharp enough to perform that minor surgery. This choice caused a rift between the two doctors because Sawyer felt his position had been threatened and warned Boone not to provide medical care for Harding unless Sawyer approved the procedure.

The president hurried to complete scheduled appearances when he arrived on shore at Seattle, Washington. Harding, who did not wear a hat, rode in a parade and presented a speech while exposed to intense sunlight. No accounts mention that he possibly suffered sunstroke but, considering how his health soon worsened, a heat-related condition might have exacerbated other physical problems, including any toxin-related damage if he had been poisoned. The weakened Harding decided to cancel his trip to Portland, Oregon, and boarded the presidential train bound for California.

During the night, Harding again experienced pain in his chest and more nausea and requested Sawyer's assistance. Boone accompanied Sawyer to evaluate Harding and realized Harding's heart was functioning erratically and was dilated. Boone insisted the train speed nonstop to San Francisco, California. The U.S. Secretary of the Interior Hubert Work, who also had a medical degree, witnessed Harding's physical distress.

San Francisco

When Harding reached San Francisco on July 29, Mayor James Rolph, Jr., welcomed him at the railroad terminal. Journalists covered Harding's arrival, the president waving to greet the crowd. Boone had wanted an ambulance to meet the train, but Harding refused, choosing to walk, instead of using a wheelchair, to the limousine, and the driver was assigned to take him to his local accommodations at the Palace Hotel. He climbed steps to enter the hotel, with one reporter photographing Harding in what would be the final image of the president while he was alive. Harding rested in his suite. His discomfort intensified during the afternoon, necessitating cancellation of plans for public appearances.

At a press conference that day, Sawyer announced Harding had been sickened by tainted seafood, attributing the president's illness to ptomaine poisoning involving harmful bacteria Sawyer believed had been transmitted from infected crabs. He assured reporters assembled at the hotel that Harding would recover, insisting the president just needed to sleep and relax. Harding's physicians released bulletins with medical information about Harding. They emphasized the bulletins would provide sufficient details about Harding to prevent speculation

like that which had occurred when President Woodrow Wilson had suffered a publicly unidentified sickness in the fall of 1919.

Boone and Work contacted Ray Lyman Wilbur, who had served as the American Medical Association president and was the current Stanford University president, and he agreed to examine Harding while the president was in California. San Francisco physician Charles Minor Cooper, who specialized in cardiovascular medicine, also evaluated Harding. On July 30, despite assurances for disclosure, those doctors refused to specify what ailment had weakened Harding. Their secrecy caused journalists and the public to speculate about what might have caused the president's health to deteriorate quickly. Local journalists reported in the *San Francisco Examiner* and *Chronicle* that the ptomaine poisoning Sawyer had diagnosed Harding with had possibly provoked stomach and cardiac complications. The medical enigma surrounding Harding fueled theories and rumors, which provided the basis for later assertions that Harding had been murdered.

Official statements next reported Harding had developed pneumonia and suffered chest pains, high pulse rates, and cardiac problems. On July 31, the *New York Times* stated Harding's health had declined. That periodical said Harding's physicians had described his condition as grave and toxic. Other U.S. newspapers printed similar accounts describing Harding's poor health and his doctors' evasiveness in elaborating about his sickness. Later that day, despite the morning articles saying his condition was dire, Harding seemed better and read newspaper reports regarding his health. Instead of upsetting him, the newspapers' suggestions that he was extremely ill bemused Harding.

The next day, Sawyer reported Harding seemed to be recovering, stating the president no longer was feverish nor had congested lungs. Sawyer stressed that Harding's physical condition remained weakened and required extended rest for his recovery. Several prominent citizens offered their luxurious California properties as retreats for Harding's recuperation. Secretary of Commerce Herbert Hoover secured a site near the Pacific Ocean where Harding could relax. The August 1 *San Francisco Bulletin* noted Harding's vital signs were returning to normal and the *San Francisco Examiner* commented that Sawyer had left the president's suite for an hour-long drive, remarking to journalists that Harding was regaining his health and had consumed some food and milk.

August 2, 1923

Accounts of the night Harding died have reiterated basic facts regarding his demise but have presented varied and often vague references such as the time of his death. Many of those details contradicted other information and could not be verified, enabling speculation about what really happened when Harding died and who witnessed his death.

Although Harding remained in bed on August 2, he seemed stronger. The morning bulletin stated Harding was comfortable and would recover. Harding ate two eggs and talked with George Christian, Jr., his executive secretary, before Christian departed to deliver a speech for Harding in Los Angeles. During the day, Harding interacted with several visitors, including his sister, Charity Remsberg, who lived in Santa Ana in southern California. He remarked to her that he was exhausted. Two nurses, Sue Dauser and Ruth Powderly, who had worked in the White House, monitored Harding's vital signs. When Secretary of the Interior Work talked with Harding, the president commented he had survived his health crisis and was eager to continue his journey. That afternoon, his physicians agreed that Harding seemed to be stabilizing.

Most accounts concur that early in the evening the president's wife entertained him by reading aloud. She stopped while sharing a *Dearborn Independent* article about Henry Ford to fluff the pillows on which Harding was braced. Harding asked her to continue reading before a convulsion abruptly shook his body. Seeking help, Florence Harding left the suite to alert Sawyer in his nearby room. Harding's physicians hurried to assist the president, but he had already died when the men reached him.

Warren Harding served as president of the United States from 1921 to 1923. Though Harding promised a return to prosperity in his campaign of 1920, his administration was plagued by scandals. He died three years into his term, allegedly of food poisoning. (Library of Congress)

Some versions stated other people, including physicians, were in the room when Harding began to sweat, become pale, and experience spasms. Witnesses saw Florence Harding insert a finger in her husband's mouth, which she said she had done to prevent him from choking on chewing gum. Sawyer took Harding's pulse, and Nurse Powderly removed Harding's pajamas, which were soaked in sweat, and toweled him dry before redressing him. Agitated, Harding expressed confusion about the unpleasant physical sensations in his body, then calmed. Because Harding's pulse was steadier, Boone left for a break while Sawyer and the nurses remained with Harding. Soon thereafter, Harding died. The official time of death varied from 7:20 to 7:35 P.M., causing many people to wonder why presidential staff could not specify exactly when Harding died and if they were hiding information.

A bulletin distributed to reporters in the hotel that night announced Harding's death. Most journalists were unprepared for that news, expecting instead to be updated on the president's recovery, which Sawyer had said was imminent. Many local reporters for San Francisco publications had made other plans that night because, accepting Sawyer's declaration that Harding was healing, they did not think there would be any breaking news about Harding to report. Stephen Early, representing the Associated Press, had been in the hall outside the presidential suite when Harding died. He telegraphed his report of Harding's death, and, during the night, the news spread across the nation and internationally.

Presidential staff began planning Harding's funeral. Florence Harding did not allow officials to conduct an autopsy of her husband. No known tissue, blood, or hair samples were preserved. She also denied casting of a death mask, a procedure traditionally performed for deceased U.S. presidents. Morticians from the N. Gray & Company started embalming Harding's body an hour after he died, preparing him for transportation within 24 hours of his death. Florence Harding privately viewed her husband in his coffin in the early afternoon on August 3. The funeral train then headed east, passing by several million mourners, to Washington, D.C., for a memorial ceremony before taking Harding to Ohio where he was buried in his hometown cemetery on August 10.

Medical Conflicts

The five physicians attending Harding in San Francisco signed his death certificate. They stated the president had died from cerebral apoplexy when a blood vessel burst in his brain after he had suffered a severe infection of his gastrointestinal tract and pneumonia. They noted Harding had also suffered from arteriosclerosis, which interfered with his circulatory and brain functions. The August 18, 1923, issue of the *Journal of the American Medical Association* printed excerpts from the physicians' medical bulletins from July 30 through August 2, concluding with several paragraphs describing Harding's health history and condition the day he died. Wilbur published an article in the October 13, 1923, issue of the *Saturday Evening Post* that reiterated the suddenness of Harding's death, especially because the president's pulse and temperature had remained near normal levels the day before he died.

Some people have hinted Harding died as a result of negligence and unprofessional medical treatment, which could be characterized as malpractice. Because Sawyer was a homeopath, many scholars have discredited his medical care of Harding. Sawyer might have unintentionally killed Harding due to his reliance on alternative medical practices and rivalry with other doctors. By forcing the other physicians, whom he perceived as competitors, to assume subordinate roles, Sawyer egotistically protected his status as Harding's main doctor in San Francisco and possibly ignored any suggestions the president suffered from

coronary disease and needed applicable medical care available at that time for that condition.

Boone later commented that he and the other physicians who had examined Harding in San Francisco considered Sawyer's methods contrary to normal techniques. Sawyer's inadequate treatment relied on purgatives for indigestion and ptomaine poisoning instead of recognizing Harding was suffering heart-related problems. Sawyer focused on purging poisons from Harding's body, which became traumatized. Patients sometimes lose potassium, which affects heart rhythm, when given too many purgatives. Contemporaries referred to Sawyer exhibiting a physical and mental decline as he aged, which may have affected his ability to practice medicine. Sawyer did not correlate Harding's declining strength with his use of purgatives, which might have stressed the president's cardiovascular functions and provoked a heart attack. If he had been aware of his shortcomings, Sawyer might have urged Florence Harding to forbid an autopsy so his negligence would remain secret.

Contemporary Reactions

Soon after Harding died some people suggested he might have killed himself. The suicide theory blamed such scandals as Teapot Dome and the betrayal by corrupt friends, including Director of Veterans Affairs Charles Forbes, for provoking a despairing Harding to commit suicide. Supporters of this theory noted Harding had revised his will prior to the western trip.

Popular culture incorporated Harding death theories. A 1926 novel, *Revelry*, written by journalist Samuel Hopkins Adams, depicted a suicidal president. Adams created characters who shared similar traits as people in Harding's administration, particularly the fictional president, Willis Markham. Like Harding, Markham has a mistress and hides her sleeping pills in his medicine cabinet because he is afraid she might overdose. Sleepless, he grabs what he thinks are sleeping pills one night but instead accidentally swallows poison. Although Markham summons his physician, he decides not to reveal he was poisoned and request medical intervention that could save his life, preferring to die and not be confronted by various scandals that will damage his presidency. Adams's book became a bestseller with 100,000 copies sold and was adapted for dramatic performances both on stage and film, spreading the suicide theory.

Adams said he was aware of people familiar with the Hardings, including a respected judge who believed Florence Harding had killed the president. Rumors circulated about Florence Harding's possible motivations for poisoning her husband. Reporters analyzed her demeanor when Harding died, contemplating why she had remained calm despite her sudden loss. Although some people perceived her behavior as being courageous, others interpreted her response as that of an uncaring spouse. Contemporaries described Florence Harding as

being a forceful, assertive person accustomed to receiving whatever she demanded and pursuing her goals with minimal regard of how she impacted others. Many people thought her actions were suspicious when she burned many of her husband's papers and asked correspondents to return letters in order to prevent them from becoming public. They concluded she was attempting to hide evidence by refusing an autopsy that might have revealed poison in Harding's system and a death mask that might have shown facial contortions caused by poison.

Some people believed Florence Harding and Sawyer might have conspired to kill the president. As the First Lady's accomplice, Sawyer might have purposefully mistreated Harding. The year following her husband's death, Florence Harding visited Sawyer's residence, White Oaks Farm, in Marion. He died at home on September 23, 1924, while Mrs. Harding was there. His family was surprised by his unexpected death with symptoms resembling those Warren Harding experienced when he died. Speculation that Florence Harding might have poisoned Sawyer reinforced many people's belief she was guilty of murdering her husband. When Florence Harding died from kidney disease on November 21, 1924, numerous facts regarding her husband's and Sawyer's deaths remained unclear.

The Strange Death of President Harding

Gaston B. Means, who had been employed as a U.S. Department of Justice investigator, arranged in summer 1928 for ghostwriter May Dixon Thacker to write a book titled *The Strange Death of President Harding*, which was published two years later and sold 80,000 copies soon after it was released. In early conversations with Thacker, Means emphasized his insider knowledge of the Harding presidency. He stressed he was aware of details regarding Harding's affair with Nan Britton and their daughter. Means also said he had information concerning premature deaths of people associated with the Harding administration. Most significantly, Means asserted Harding's wife killed the president with poison.

Means's book described how Florence Harding spent time alone with her husband the night he died. Means said the First Lady had told him that she handed her husband's medication to him and he swallowed all of it before dying. The former First Lady implied that she had poisoned that medicine. Means suggested two motives for her action. Means thought Florence Harding had murdered her husband to save him from suffering being discredited publicly for the unethical activities of his cronies. She wanted Harding's constituents to respect him, believing her husband would retain popular support among the grieving populace. Means also thought she might have poisoned Harding as vengeance for his womanizing, particularly with Britton. He declared she had expressed hatred for Harding and emphasized that her husband had to die as a

result of his various transgressions and to save the Republican Party and the United States. According to Means, Harding was not regretful of her actions and viewed them as fulfilling her destiny.

Means characterized the president's widow as ambitiously desiring power and refusing to be dishonored by her husband's affairs and the greed of his associates. He said that he and Mrs. Harding had talked at the time of President Harding's funeral. They discussed how the Ohio Gang had manipulated and exploited the president and she feared his impeachment was inevitable. She perceived herself as the sole person who could protect her husband from his enemies' control by ending his life and thus safeguarding his legacy and honor. Means claimed Mrs. Harding consulted him about preventing any autopsies of the president.

He suggested Florence Harding had schemed with Sawyer to kill the president. In an epilogue, Means stated Sawyer had said in a conversation with Means that he was not involved directly with poisoning Harding. According to Sawyer, Florence Harding excused the president's secret service guard and nurses on duty, saying she and Sawyer would tend to the president. Sawyer left the couple alone for 10 minutes, returning to find the president dead. Implying the First Lady poisoned her husband, Sawyer ambiguously referred to a flawed poisoning in Vancouver.

Murder Theories

Means's book expanded speculation nationwide in the 1930s about Harding's possible murder and convinced many readers. The year after Means's book was published, John W. Gunn published a booklet, *Was President Harding Murdered?* (n.d.), and Frederick Lewis Allen discussed Harding being poisoned in his book *Only Yesterday: An Informal History of the 1920s* (1931). Allen agreed that Florence Harding might have murdered her husband and been assisted by Sawyer. He acknowledged Harding's death could have resulted from suicide.

Many contemporary Americans wondered if the secrecy Florence Harding and her associates maintained about specifics of President Harding's death was their strategy to conceal truths about why he died. Most theories focused on poison as the cause of Harding's death. People developed variations of how Harding was poisoned. One theory speculated Florence Harding had initially attempted to poison her husband when they were at Vancouver, an idea compatible with Means's reference. The amount she gave him was insufficient to be fatal but did sicken the president. This account usually says she poisoned Harding's food and his reaction to what he consumed resulted in his physicians' assumption he had been food poisoned instead of realizing the president had been given poison.

After several days in San Francisco, Harding seemed healthier because his system had rid itself of the poison his wife had given him. Theorists thought

Florence Harding took advantage of the time alone with her husband to give him a stronger, deadly dose of poison. She might have rubbed poison on his gums when she put a finger in his mouth. Because Sawyer injected Harding with digitalis, Mrs. Harding, or another poisoner, might have slipped a toxin in the hypodermic needle. Some people who endorsed this theory believe Sawyer was Florence Harding's accomplice or later realized she had murdered her husband, forcing her to poison him too so he could not incriminate her. In his 1939 book, *Fighting Years*, prominent editor Oswald Garrison Villard stated he believed Harding probably had been murdered, commenting about suspicious coincidences regarding Sawyer's death.

Other theories circulated. A cook claimed he had put croton oil on Harding's food. Some people suggested the Ku Klux Klan had ordered members to kill Harding because he possibly had African American ancestry. Assassins theorists identified included Madame Marcia whom people thought might have murdered Harding to preserve her credibility for her prediction he would die in an unusual manner. Several poisoning theories targeted various government officials as potential villains. Some plots accused that corrupt politicians killed Harding so he could not testify against them or threaten their continued pursuit of unlawful activities involving government resources.

Walter Thayer, who had been a secret service agent when Harding was president, collected evidence for a comprehensive, three-volume report titled "The Harding Poison Murder Case." Thayer theorized Harding had first been poisoned en route to Alaska. The president was poisoned again just prior to reaching Seattle. A third poisoning of Harding occurred on the train to San Francisco where he received a fourth dose of poison that killed him. Thayer stated the poisoners were German agents who acted on behalf of a Harding associate who wanted the president dead. He outlined how each poisoning coincided with the president's health lapses during the trip.

Views that Harding had been murdered persisted in the decades after his death. Technological advances in the 20th century extended debate whether Harding had been murdered. Radio and television broadcasts offered forums for people to explain their opinions about Harding's death. C-Span aired its *American Presidents: Life Portraits* episode about Harding from Marion, Ohio, in September 1999. Many of the viewers' questions regarded whether Harding had been murdered. The Internet provided people, primarily nonhistorians, a new medium to discuss their theories about the causes of Harding's death. People posted ideas about possible murderers and motives on Web sites. Some posters suggested organized crime figures had assassinated Harding. A few individuals claimed personal associations with Harding, such as a woman who said she was his relative and had genealogical evidence proving Harding had an African American ancestor, a reason she emphasized for some people to want Harding killed.

Conclusion

The precise reason Harding died will remain unknown because an autopsy was not conducted immediately after his demise to clarify exactly what caused his death. No medical officials at the scene considered looking for poison in his system or saving tissue samples. Later assessments of contemporary medical reports by physicians suggested possible causes, but their hypotheses cannot be proven without physical evidence. Any 21st century exhumation of Harding's remains might not offer closure regarding toxins he might have ingested.

Rumors that Florence Harding killed her husband endured. Almost two decades after Harding's death, Adams encountered Ohioans, many of whom had known the Hardings and believed she was guilty. If she had not maliciously poisoned her husband, Florence Harding was so self-absorbed that she might have accidentally given him her medicine, including codeine, which might have contradicted or created a toxic mixture with his medications. Some theorists stressed that many homeopathic remedies involved traces of toxic substances, such as the morphine Harding's mother had mixed in medications, suggesting Harding might have built up sufficient levels of poison during his trip. They noted the symptoms Harding experienced were occasionally connected to poisons. Without an autopsy, those poisons could not be confirmed to be present in Harding's body.

Surrounded by strangers and friends, Harding was vulnerable to attack. Unavailable or conflicting information about people with whom he had contact contributed to murder theories. For example, some accounts named people who prepared Harding's meals and had access to his food, including Phillip Roemer, the Palace Hotel head chef, in addition to vague references to anonymous kitchen help. Several contemporary articles commented on physical stresses often associated with the demands of the presidency. Perhaps one of Harding's enemies took advantage of his weakened physical and emotional condition due to the toils of traveling, corruption revelations, and family history of poor cardiovascular health. Recognizing Harding's vulnerability, the murderer, whether a relative, associate, or stranger, inflicted a fatal dose of poison that overwhelmed Harding.

References and Further Reading

Allen, Frederick Lewis. *Only Yesterday: An Informal History of the 1920s*. New York: Harper & Bros., 1931.

Anthony, Carl Sferrazza. *Florence Harding: The First Lady, the Jazz Age, and the Death of America's Most Scandalous President*. New York: Morrow, 1998.

Bumgarner, John R. *The Health of the Presidents: The 41 United States Presidents through 1993 from a Physician's Point of View*. Foreword by Walter L. Floyd. Jefferson, NC: McFarland, 1994.

Dean, John W. *Warren G. Harding*. New York: Times Books/Henry Holt, 2004.

Deppisch, Ludwig M. *The White House Physician: A History from Washington to George W. Bush*. Jefferson, NC: McFarland, 2007.

Ferrell, Robert H. *The Strange Deaths of President Harding*. Columbia: University of Missouri Press, 1996.

Ficken, Robert E. "President Harding Visits Seattle." *Pacific Northwest Quarterly* 66, 3 (July 1975): 105–14.

Gunn, John W. *Was President Harding Murdered?* Girard, KS: Haldeman-Julius Company, n.d.

MacMahon, Edward B., and Leonard Curry. *Medical Cover-Ups in the White House*. Washington, DC: Farragut, 1987.

Payne, Phillip G. *Dead Last: The Public Memory of Warren G. Harding's Scandalous Legacy*. Athens: Ohio University Press, 2009.

Schruben, Francis W. "An Even Stranger Death of President Harding." *Southern California Quarterly* 48 (March 1966): 57–84.

Whitaker, W. Richard. "The Night Harding Died." *Journalism History* 1, 1 (Spring 1974): 16–19.

CON

Shortly after the funeral, questions arose and conspiracy buffs started their theorizing. The funeral in question was that of Warren G. Harding, 29th president of the United States; the date was August 2, 1923; and the place was San Francisco's Palace Hotel.

Without doubt, the situation surrounding Harding's death led to questions. How had such a seemingly healthy man died so unexpectedly? With an illness obvious to the public for only a week, his passing seemed improbable. Had the president actually died from ptomaine poisoning from ingesting spoiled crabs? Or, more menacingly, had the president been murdered? Tensions had been high between him and his wife—Florence Kling DeWolfe Harding—almost from the day of their wedding. Had she finally had enough of his philandering? Rumors of his mistress Nan Britton and their infant daughter Elizabeth Ann, born October 22, 1919, had been circling the White House for years. Tales of other trysts abounded. Had her humiliation finally reached the point of no return?

Or, as others soon would have it, had the new widow poisoned her beloved "Wurr'n" to save him the humiliation of the scandals that would soon become known to all? Beyond the extramarital affairs, it was soon to become common knowledge that Harding had unwittingly appointed and then overlooked the actions of numerous corrupt officials. Secretary of the Interior Albert Fall sold the nation's oil reserves at Wyoming's Teapot Dome for his own personal profit. Charles

Forbes, corrupt director of the Veterans Bureau, callously diverted drugs from veterans' hospitals, skimmed the proceeds from sales of "surplus" goods, and gained considerable kickbacks from the purchase of supplies at exorbitant prices. Thomas Miller, the alien property custodian, accepted bribes with alacrity, ultimately earning himself a prison sentence. That Harding himself profited from these machinations has never been proven, and it appears that he learned only weeks before his death of the activities of Fall and Forbes. Had the First Lady poisoned the president in order to save him the ignominy of a failed presidency, possible impeachment, or the denial of the second term that he so desperately sought?

Others muttered about the possibility of the president committing suicide. Similar to those who theorized that Florence Harding had murdered her husband, suicide conspirators noted that the president had undoubtedly known that his presidency was about to collapse. Harding had proven himself a notoriously poor judge of character, appointing to office cronies from all-night poker parties and shallow friendships. His faith in "friends" generally proved in error, with the ultimate burden of their unscrupulous activities falling upon his shoulders.

With regard to Forbes's malfeasance, one estimate has his activities costing the taxpayers an approximate $200 million. That Harding was at least partially aware of Forbes's misbehavior is unmistakable. According to one witness, Harding ultimately confronted Forbes in the Oval Office, shaking, or alternatively strangling him, yelling, "You yellow rat! You double-crossing bastard!" (Trani and Wilson, 1977: 182). Indeed, not long after learning of Forbes's betrayal, Harding told Kansas newspaperman William Allen White, "My God, this is a hell of a job! I have no trouble with my enemies; I can take care of my enemies alright. But my damned friends . . . my God-damn friends, White, they're the ones that keep me walking the floor nights!" (Russell 1968: 558, 560).

That his soul was troubled about his friends and their antics became clear in a conversation with Secretary of Commerce, and future president, Herbert Hoover, as they traveled to Alaska on Harding's final trip. As Hoover recalled nearly 30 years later:

> [O]ne day after lunch, when we were a few days out, Harding asked me to come to his cabin. He plumped at me the question: "If you knew of a great scandal in our administration, would you for the good of the country and the party expose in publicly or would you bury it?" My natural reply was, "Publish it, and at least get credit for integrity on your side." He remarked that this method might be politically dangerous. I asked for more particulars. He said that he had received some rumors of irregularities centering around [Jess] Smith, in connection with cases in the Department of Justice. . . . Harding gave me no information about what Smith had been up to. I asked what [Attorney General Harry] Daugherty's relations to the affair were. He abruptly dried up and never raised the question again. (Hoover 1952: II-49)

Clearly, matters lay heavy upon the president's mind, not only about the present, but also about the future of the Republican Party and his own personal fortunes.

Each of the theories gained adherents over the course of the decade following Harding's death. The suicide theory gained popularity with the publication of Samuel Hopkins Adams's *Revelry*, a purported novel that examined the death of President Willis Markham. Markham, innocent of the actions of his cabinet members and appointees and fearful of the nation's retribution upon him and his beloved Republican Party, frets about a fictionalized Teapot Dome scandal. Only by accident, however, does Markham take the wrong and fatal prescription. Realizing his error, Markham chooses to do nothing, ultimately succumbing to the medication and ending his grief over false friends.

Gaston Means's *The Strange Death of President Harding* placed the burden of the medicine chest squarely on the First Lady. According to the scurrilous work, Florence Harding had deliberately poisoned her husband. As to precisely why, Means was a bit unclear. She had poisoned him either to protect him from the forthcoming announcements of ill behavior by cabinet members and political appointees or else in retribution for his long-term affair with Nan Britton and their resultant child.

Or was it as pure and simple as accidental food poisoning incurred from crabs, fresh or canned, that led to the president's demise? That was the initial call by his homeopathic physician Charles Sawyer.

The President's Health

Despite a rather robust appearance—six feet tall, large-boned, and full-chested—Warren Harding could not claim the perfect health of a corn-fed Ohio country boy. Rather his health had always been somewhat delicate. In addition to the youthful maladies of mumps, chickenpox, and measles, Harding's constitution was beleaguered by frequent heartburn and indigestion.

Frazzled nerves also seemed to plague the future president. Indeed, while a young man of 24, Harding suffered a nervous breakdown, unable to attend to his work at the *Marion* [Ohio] *Star*. On November 7, 1889, he checked himself into the Battle Creek Sanitarium in Michigan. Sponsored by the Seventh-day Adventist Church, the vegetarian facility was run by J. P. Kellogg, inventor of corn flakes. In early 1890, Harding returned to Marion, 20 pounds lighter and with spirits significantly lifted.

His spirits, however, did not remain high, nor did his general health. Not too long after his marriage to Florence Kling DeWolfe on July 8, 1891, the Hardings moved in with his parents, both of whom were homeopathic physicians. For nearly six months, the newlyweds lived with his parents in order to spare his father late night trips to the younger Hardings' home to treat Warren's heartburn

and sudden attacks of indigestion. And, on January 7, 1895, Harding again checked himself into the Battle Creek Sanitarium. He remained there until late February, briefly returned home, and then went back to the sanitarium for much of the remaining year. It appears he would regain his health under the Spartan diet and quiet of the sanitarium, only to again lose his composure upon returning home to a life of late nights, unabashed drinking, a poor diet, constant cigar smoking, and a perpetually bothersome wife. In 1896 and 1897, he again paid brief visits to the sanatorium, as he did for a few days in 1903.

Harding's discontinuation of Kellogg's treatments at the Battle Creek facility did not mean a return to health. Rather, he gathered his emotions, continued eating and drinking to excess, smoking and staying up until the wee hours of the morning, suffering heartburn and indigestion along the way. He also found comfort in the arms of at least two mistresses—Carrie Phillips and Nan Britton. The tryst with Phillips lasted from 1905 through 1919 and the one with Britton lasting from 1917 until shortly before the president's death.

Turning his health over to his father and a family friend, Charles Sawyer, a homeopathic physician, Harding remained relatively healthy over the course of the first two decades of the 20th century. In 1919, however, he told a senate confidant that he suspected he had some sort of heart ailment. Also in 1919, he noted to the same confidant—Senator James E. Watson of Indiana—that his urine contained traces of sugar and that his systolic blood pressure was occasionally as high as 175.

Throughout the campaign of 1920, Harding appeared hale and healthy, if not occasionally exhausted by the strenuous nature of presidential politics. By 1922, however, the strains of office were beginning to show to all who cared to notice. He was much more easily exhausted, complained of frequent chest pain, and often had trouble sleeping. While Sawyer was willing to suggest that all could be traced back to the president's notorious attacks of indigestion, both patient and physician suspected that the source of Harding's ill health lay in angina pectoris: acute chest pain or discomfort resulting from decreased blood supply to the heart muscle.

Outsiders who noticed the president's haggard look generally assumed that worry over the health of his wife—who had a severe kidney ailment—or the recent strikes in the railroad and coal industries weighed heavily upon him. While they may indeed have weighed heavily upon him, so was his own deteriorating health. His chest pains continued and his systolic blood pressure soared into the 180s. Still he did not sharply curtail his schedule, either socially or politically, as he was seriously considering a second term of office. During a social event, the president made the passing acquaintance of Emmanuel Libmann, a noted New York cardiologist. Libmann observed the president over the course of the evening, noting a frequent shortness of breath, as well as abrupt pauses in his conversation. The next day, Libmann predicted that the

Funeral cortege of the late President Warren Harding leaving the White House, August 8, 1923. (Library of Congress)

president would be dead within six months. He was off by less than two months in his estimation.

A serious attack of influenza laid the president low in January 1923 and *may* have been accompanied by a mild heart attack. His recovery was slow, the impact long-lasting and noticeable to all. His valet Arthur Brooks confided to Colonel Edmund Starling of the secret service details that something was seriously wrong with the president. He could not sleep at night, nor could he lay flat. He could only drowse fitfully if propped up by pillows. Even then he had trouble catching his breath. Starling shared with Brooks that the president could no longer play a full round of golf, usually tiring out at the halfway point in the round.

Over the course of the spring, the president's health continued to deteriorate. His energy was nonexistent; his usually ruddy color had turned an ashen gray. He complained to Secretary of State Charles Evan Hughes that not only was he perpetually exhausted, but that his blood pressure consistently maintained 175, systolic. The time hardly seemed appropriate to commence a nationwide tour, but this is exactly what the president and his handlers wanted in light of the forthcoming election: a time to build support, a time to meet the common laborer and farmer, a chance to explain decisions that had been made and to test the waters on ones that would be made. Thus the "Voyage of Understanding" came into being and the president's life became more and more difficult to sustain.

The Voyage of Understanding

The president's cross-country trip was not only to gauge the American public, but also to reinvigorate the exhausted man, restoring his health in the process. Scheduled to last slightly more than two months, the trip was to take Harding across the western United States and on to Alaska, returning to Washington, D.C., via the Panama Canal. While a two-month vacation from the presidency might shock modern sensibilities, the presidency was a much less intensive position in the early 20th century than it is today, and Harding had appointed quality cabinet members for the most part to carry on the day-to-day activities of his office. In the sultry, pre–air conditioning days of Washington, D.C., most individuals with the wherewithal to leave did so.

Prior to leaving on his lengthy tour of the western states, Harding arranged for Postmaster General Walter Brown to travel his proposed route and arrange for speaking engagements and assorted activities. When Brown returned from his foray, Harding complained to Colonel Edmund Starling of the secret service that the trip would be exhausting, just the thing he was hoping to avoid. Harding needed rest, and he knew it. Accordingly, Starling trimmed the schedule considerably, but not enough for the life of the president to be saved.

On June 20, 1923, President Harding, the First Lady, Secretary of Agriculture Henry C. Wallace and his wife, Secretary of the Interior Hubert Work and his wife, Speaker of the House Frederick H. Gillett, Charles Sawyer and his wife, and a contingent of newspapermen, secret service men, aides, stenographers, and telephone technicians, meant to arrange for the amplification and transmission of the president's speeches, departed the capital. Secretary of Commerce Herbert Hoover and his wife were to join the party in Tacoma, Washington. Also joining the party was Joel T. Boone, a homeopathic physician meant to shore up the elderly and increasingly infirm Sawyer.

As the date for departure grew near, Florence Harding assumed a poise of concern for the president's health. She approached Starling and demanded that Sawyer and Boone be placed as closely to the president as possible at each stop. Adjoining rooms would be preferable, but if not available, the First Lady wanted to know the exact room number for each of the physicians, in case she might need them in an emergency.

Regardless of Starling's merciless paring of the schedule, the trip was still exhausting, even to the younger, healthier members of the party. The *Superb*, the Hardings' special railcar, stopped on a relentless schedule, exhausting the president, sallowing his complexion and dulling his senses. From West Virginia, through Ohio, Indiana and Illinois, into the searing heat of summer, Harding appeared time and again at whistle stop after whistle stop. In the process, his lips grew blistered, requiring Sawyer to apply ice compresses to them. From Saint Louis on to Kansas City, Missouri, the president continued his constant

speech-making, essentially doing nothing to ease the strain on his heart. William Allen White, a Kansas editor, met with Harding in Kansas City. White later recalled that Harding's lips were swollen and blue, his eyes puffed, and his handshake stiff.

As Harding continued across the nation, he kept a blistering pace: speaking, touring the site of the future Zion National Park, visiting soldiers in veterans' hospitals, and even taking the time to take the controls of a Kansas wheat binder for a few moments of farm labor. He also continued his "Voyage of Understanding" for the American people by speaking on the World Court, the 18th Amendment, railroad and transportation problems, taxation, law enforcement, social justice and labor, internal improvements, immigration, and Americanization. Each exhausting stop was calculated to explain to America what his administration had accomplished to date and what it hoped to accomplish if returned to office in 1924.

The trip west had taken a serious toll upon the president. He was not sleeping well and when he did manage a few minutes' sleep, it had to be taken sitting up. His complexion remained sallow while his breathing was shallow and somewhat rapid. Still, he had taken every opportunity possible to speak to the American people and was looking forward to the brief respite of sea travel to Alaska on the USS *Henderson*.

The trip to Alaska, a mere four days in duration, did provide some relief to the exhausted president. There were no speeches to deliver, no hands to shake, no feigned interest to be taken, no meaningless side trips to be endured. He did, however, exhaust his traveling companions. Unable to sleep, he played cards into the wee hours of the morning, much to the distress of the few bridge players among the party. His worsening heart, the excitement of the journey, and lengthening days kept him restless; he paced the deck if not playing cards.

At an inland stop in Fairbanks, Harding endured 96 degree temperatures as he delivered his prepared remarks. However, the fatigue etched in his face distressed the normally unastute Sawyer to such a degree as to recommend that the president take a direct rail route back to the *Henderson* rather than the planned-for auto trip via the rugged Richardson Trail. He felt the president needed at least two good days' rest.

Once aboard the *Henderson* and southward bound, Harding's restlessness continued. He paced the deck nervously, showed signs of moroseness, and could not sleep without the assistance of heavy sedatives. And when sedatives could not induce sleep, Harding again dragooned bridge players for late night/early morning sessions, exhausting his compatriots and forever ending Hoover's love of the game.

One night in a fit of restlessness, Harding wandered the decks of the *Henderson*. In his meandering, he came across Reddy Baldinger, former senate page for Harding and at one point a delivery boy for the *Marion Star*, who was

eating a pile of fresh crabs. Baldinger offered to share his feast and the two men sat there in the night, cracking shells, digging out small bits of meat, letting butter drizzle down their chins. Gifts of crabs and other seafood came freely to the *Henderson*. On one occasion, not only did Harding become ill, but so did several members of the presidential party. These late-night meals of crabs would be blamed by some for the president's death via food poisoning.

Yet even before the crabs had digested, the press corps was noticing that all did not seem well with the president. One noted that the president was "not just tired or worn out. He is beyond being merely fatigued. He is an entirely exhausted man, and the sort of rest he requires is not that of a day, or two or three, but of weeks" (Murray 1969: 447). Although Sawyer insisted that a serious case of indigestion was to blame, Boone and Work had done a quick examination of the president and declared his heart to be alarmingly enlarged.

A stop at Vancouver revealed greater concerns about the president's health. While he manfully delivered a speech to 40,000 individuals gathered in Stanley Park, bareheaded, many of his party noticed his exhausted demeanor and near frail appearance. Later that day on a golf course, the president could not finish his round, playing the first through sixth holes and then quietly slipping to the seventeenth to finish out his round. A formal dinner found the president listless, delivering a forceless 15-minute speech. Concern for the president's health grew as he sought out every available moment to lie down and rest.

As the presidential entourage traveled into U.S. waters, Harding's health continued to deteriorate. When fog kept the *Henderson* from docking until several hours after its scheduled arrival, all believed that the day's events would be foreshortened. However, rather than simply starting at noon and forgoing the morning's scheduled activities, the president was kept active all day and into the evening hours. In mid-afternoon, Harding delivered what was to be a major speech about the Alaskan territory. However, he was listless, called Alaska Nebraska, and occasionally allowed his voice to fall to a mere whisper. He even dropped his speech, forcing Hoover to reassemble it on the spot. His traveling companions clearly understood that the president was ill.

While Sawyer continued to assert that the president's illness could be blamed on spoiled crabs that had induced acute gastrointestinal distress, his other physicians were not so inclined. In a quiet moment when Boone was alone with Harding—Sawyer was jealous of the Hardings' burgeoning friendship with Boone— he took the opportunity to examine the exhausted man. Along with Secretary of the Interior Work, Boone took the president's vital signs: his pulse raced at 140 beats per minute, his respiration had risen to 40 breaths per minute, and his heart seemed grossly enlarged. In consultation with Secretary Hoover, the physicians determined to contact one of California's leading doctors: Ray Lyman Wilbur, then the current president of Stanford University and soon to be president of the American Medical Association. He was to bring a heart specialist with him. He

selected Charles M. Cooper. All stops between Seattle and San Francisco were cancelled as the patient sped south toward his physicians.

By July 29, 1923, a Sunday, Harding's condition had worsened. Still the president insisted that he would not be seen in a wheelchair or pajamas and accordingly dressed and walked from the station to a waiting car. Arriving at the Palace Hotel in the early part of the day, Harding was immediately sent to bed. It was abundantly clear to all, save Sawyer, that the president's condition did not derive from tainted crab, but rather resulted from coronary distress. Blood tests and x-rays revealed that beyond his heart woes, he was also suffering from pneumonia. With a pulse rate of 120 and a temperature of 102, things did not look well at all. Tuesday and Wednesday seemed to show improvement, so much so that Sawyer announced that he himself would soon be taking a vacation. On Thursday morning, the generally pessimistic Boone was even optimistic that the presidential party could soon leave California for Washington, D.C., and at least two months of enforced rest.

Although Boone, Wilbur, and Cooper attempted to stimulate the president's body via doses of digitalis and caffeine, Sawyer, the senior and official White House physician, continued his own personal round of treatments. Attempting to purge the president's body of the ptomaine poisoning he still believed to be the primary difficulty, Sawyer applied a variety of purgatives to his patient. Regardless of the assorted treatments, Harding seemed to be on the mend by Thursday. He sat up in bed and visited with a number of members from his entourage.

Early in the evening of August 2, 1923, the First Lady commenced to read an article to the president. From the pages of the *Saturday Evening Post*, she lifted his spirits with Samuel Blythe's "A Calm View of a Calm Man," an essay that lauded the president's steady, competent course of action. During her reading, the president suffered a seizure of sorts when he noticeably stiffened, showed untoward signs of irritability, began sweating profusely, and developed a very frightened look. Once he was calmed and his pajamas were changed for dry ones, Florence crossed the hallway to her own bedroom to rest. Sawyer stayed with the president for some time, most probably administering yet one more purgative to ease the poisons from his body.

What happened next is the material of conspiracy theories. Who was in the room at the time of Warren Harding's death remains a mystery. Was the First Lady even present to see her husband draw his final breath? The number who claim to have seen the president twist convulsively in death and then fall lax varies with the telling, as does the time of death. Did he die at 7:10, 7:20, or 7:30? Even the exact cause of his death is still debated by historians and conspiracy theorists. Was it ptomaine poisoning from bad crabs? Was it poison from an alternate source? Was it a heart attack? Was it a cerebral hemorrhage? And why did the First Lady refuse to allow an autopsy of her comparatively young—57 years of age—husband?

The Harding Memorial

Following the death of Warren Harding, Vice President Calvin Coolidge was staying with his father at Plymouth Notch, Vermont. In the sitting room of the family house, Coolidge was sworn in as president by his father, a notary public.

Harding's body was brought by train from San Francisco to Washington, D.C., and his body was first taken to the East Room of the White House, and then lay in state on the same catafalque that had been used for Abraham Lincoln in 1865. Some 30,000 people filed past his body.

Then Harding's body was taken to Marion, Ohio, where his father lived. It was placed in a temporary vault at Marion Cemetery, as was that of his wife, Florence, who died in the following year. In 1931 the Harding Memorial, a large tomb, was formally dedicated at Marion for the Hardings. Herbert Hoover presided over the service. It remains open to the public.

Credible Conclusions

In the nearly 80 years since Harding's death, a number of volumes have been produced by historians exploring his presidential administration, as well as those that attempt to present a picture of his entire life. Since the president's papers became available for public scholarship in the 1960s, no academician has made a credible case for food poisoning or, for that matter, poisoning of any kind. Indeed, only Robert Farrell, in his *The Strange Deaths of President Harding* (1996), spends any pages on the poisoning or murder theories, and then only to debunk them. Francis Russell (*In the Shadow of Blooming Grove: Warren G. Harding and His Times*, 1968) passes over the controversies rather rapidly, noting only that the physicians present did not agree on a precise cause of death. Robert K. Murray (*The Harding Era: Warren G. Harding and His Administration*, 1969) is similarly quiet on the controversies surrounding the death of the president.

Only in recent years has a new hypothesis been proposed relating to the death of Harding. Carl Sferrazza Anthony in *Florence Harding: The First Lady, the Jazz Age, and the Death of America's Most Scandalous President* (1998) offers the first reinterpretation of the tragedy in nearly three decades. And his argument is persuasive. According to Anthony's reading of the documentary evidence, Florence Harding and Sawyer returned to the president's room slightly before 7:00 P.M. There they joined naval nurses Ruth Powderly and Sue Dauser, who were changing shifts. In the first minutes after 7:00, Harding passed away following a quick convulsion where his mouth dropped open and his head lolled to the right. The First Lady did not at first believe her husband had died and implored Sawyer to do something to revive her husband. Sawyer reportedly ran from the room in order to retrieve a stimulant from his room cum office.

When the stimulant did not revive the president, Florence Harding stepped into the hallway and asked that all the doctors come to the sickroom immediately. Within minutes Wilbur, Work, Cooper, Sawyer, and Boone were in room 8064. Each saw the case as hopeless, the president having been dead for several minutes. To appease Florence Harding, Boone opened the president's eyes, touched the corneas, closed the eyes once again, and turned to embrace the First Lady. The time was near 7:30 P.M.

As it came time to fill out the death certificate, Sawyer, as the president's chief physician, declared that the cause of death had been a cerebral hemorrhage. Starling, the secret service agent in charge of the scene, attempted to ascertain if the other physicians believed a hemorrhage was the cause of death. For hours he interviewed the physicians, attempting to determine if unanimity existed as to the cause of the president's death. More significantly, despite his having been the physician in charge, Sawyer refused to sign the death certificate. Did he know that the certificate was a misrepresentation of the facts and signing would be tantamount to committing perjury? Ultimately, Wilbur, a physician resident to California, signed the document, despite the fact that all but Sawyer felt that death had come as a result of cardiac arrest.

For Anthony, the death of Harding falls into the category of "negligent homicide" (Anthony 1998: 453). The First Lady pushed the president too much to complete the western tour, despite knowing he was ill. Sawyer treated the president for nonexistent food poisoning by applying assorted purgatives, which ultimately altered his potassium levels, leading to a heart attack. And, after the first attack on August 2, Sawyer continued to ply the president with purgatives, most probably leading to the second and fatal heart attack.

Moreover, despite Boone having serious reservations as to the treatment plan to restore the president's health, he did little to nothing to alter Sawyer's misguided efforts. Of course, neither did Wilbur, Work, nor Cooper, at least according to their notes and memoirs. They did apparently talk among themselves and years later with friends as to the misguided efforts of the homeopathic physician, but long after the patient and the doctor himself were dead.

In the end we are left with a dead president. His family history suggests a fatal heart attack, as does the quickness of his actual passing. Cerebral hemorrhage generally brings death within minutes; all involved said the president's death was instantaneous.

The First Lady might have been angry with the president for his many trysts, but she had lived with them for years and would have had no reason to finally draw the line and murder the man who gave her a significant station in life. Not enough of the scandals by which his presidency would be remembered were yet known. The First Lady had no reason to poison the president to save him from impeachment and disgrace.

Harding himself was an unlikely candidate for suicide. Although he was cognizant of his own mortality—preparing a will shortly before leaving on his western tour—he looked forward to his continued involvement in politics and perhaps even a second term in office. He also looked forward to traveling and to writing again for his beloved *Marion Star*.

Thus, the blame must fall on the president's lifestyle and the medical care provided by Charles Sawyer. Had he not so fervently believed that the president had ingested tainted seafood, his course of purgatives would not have been prescribed, or they would have been more rapidly curtailed. But with his already weakened heart, the purgatives forced Harding into cardiac arrest.

Those seeking to find suicide or murder in the death of President Warren G. Harding should put away their convoluted conspiracy theories and accept the truth.

References and Further Reading

Anthony, Carl Sferrazza. *Florence Harding: The First Lady, the Jazz Age, and the Death of America's Most Scandalous President*. New York: Morrow, 1998.

Ferrell, Robert H. *The Strange Deaths of President Harding*. Columbia: University of Missouri Press, 1996.

Hoover, Herbert. *The Memoirs of Herbert Hoover*. 3 vols. New York: Macmillan, 1952.

Murray, Robert K. *The Harding Era: Warren G. Harding and His Administration*. Minneapolis: University of Minnesota Press, 1969.

Russell, Francis. *In the Shadow of Blooming Grove: Warren G. Harding in His Times*. New York: McGraw-Hill, 1968.

Starling, Edmund W., with Thomas Sugrue. *Starling of the White House*. Chicago: Peoples Book Club, n.d.

Trani, Eugene P., and David L. Wilson. *The Presidency of Warren G. Harding*. Lawrence: Regents Press of Kansas, 1977.

10

Marcus Garvey was "railroaded."

PRO   Kelton R. Edmonds
CON   Tim J. Watts

PRO

Marcus Mosiah Garvey arrived in the United States from Jamaica in 1916 without significant contacts and with little money. Yet, by the early 1920s, the immigrant leader amassed the largest black movement in U.S. history. Shortly after rising to prominence, Garvey saw his organizational efforts devastated due to the combination of governmental repression, international opposition, and anti-Garvey campaigns by other African American leaders. Garvey's meteoric rise can be attributed to his charisma, organizational skills, and keen understanding of the time period. The economic, social, and political circumstances of blacks within the United States at the time of his arrival proved to be an ideal opportunity for Garvey to galvanize a mass movement. Conversely, his catastrophic and relatively quick fall exposed the collective power of his enemies as they systematically railroaded Garvey into isolation, thereby totally diminishing the effectiveness of his ideology and organizations.

Born in Saint Ann's Bay, Jamaica, on August 17, 1887, Garvey migrated to the United States in 1916 after exchanging a series of letters with iconic black leader Booker T. Washington, a man whom Garvey admired for his unprecedented efforts toward black nationalism. Unfortunately, Washington died before Garvey could meet him. Nevertheless, Garvey did not let Washington's death dissuade his vision as he quickly amassed capital and momentum in organizing one of the most successful movements of the 20th century.

As an adolescent, Garvey was able to nurture a love for reading while working on ships that sailed throughout the Caribbean and South America. While traveling aboard these ships, Garvey grew disheartened as he witnessed the deplorable living conditions of blacks throughout the Americas. He concluded that wherever he traveled, people of African descent were blatantly disrespected and systemically fixed at the bottom of the socioeconomic hierarchy.

Consequently, Garvey became dedicated to addressing the universal subjugation of people of African descent throughout the world. He constructed an organization to uplift blacks on a Pan-Africanist level, called the Universal Negro Improvement Association (UNIA). Although organized in Jamaica in 1914, the

217

UNIA flourished because of its volume and vitality in the United States. Garvey particularly appealed to millions of blacks who resided in northern urban centers in the United States, many of whom had recently migrated from the oppressive conditions of the rural Jim Crow South. While pursuing northern World War I industrial employment, blacks found major injustices in the North, sometimes as bad as the South, as they were subjected to economic discrimination, de facto segregation, and episodes of racial violence.

Garvey's timely and powerful oratorical abilities struck numerous emotional chords among African Americans. He offered African Americans a message that emphasized pride and self-esteem by celebrating the richness of African history. He appealed to African Americans with a message of economic prosperity through building and supporting black autonomous enterprises. Accordingly, Garvey offered successful examples of his economic strategy for blacks to support or immolate, as the UNIA owned factories, bakeries, a grocery store, and a noteworthy shipping fleet called the Black Star Line. Racial pride, however, was just as important as the building of "black wealth," according to Garvey. Garvey boldly asserted that racial pride for blacks centered on both understanding the rich history of their African heritage and the establishment of a glorious new empire on the African continent, which European imperialist powers almost totally controlled at the time. To implement his plan of redeeming Africa, as well as spread his economic ideals, Garvey relied on his newspaper, *The Negro World*, to promote his movement throughout the world.

Preaching total separation of blacks from all forms of white domination, his eloquent speeches and articles promised reclamation, hope, and prosperity to black people. Garvey even created the opportunity for blacks to invest financially in his movement, as he sold stock shares in the Black Star Line. Garvey originated the Black Star Line steamship fleet to transport black passengers and commodities to and from Africa, the Americas, and the Caribbean. As membership skyrocketed to around 3 million, money began to pour in to UNIA divisions nationwide.

For as many people who praised Garvey's ideology and success, however, they were superseded by powerful forces that wanted him stopped and immediately removed from any influential platform. Indeed, Garvey accumulated a long list of vehement enemies, residing both inside and outside the United States. This collection of enemies resorted to unscrupulous and inhumane tactics to ultimately subdue and ostracize Garvey, destroying his rapidly growing movement. The systematic pursuit and persecution of Garvey ultimately resulted in his five-year prison sentence. This section explores the various enemies that Garvey accumulated during his rise to prominence; then it critiques how his enemies collectively discredited and railroaded Garvey, culminating in his notorious trial and conviction, which eventually led to the death of his movement.

Considering the magnitude of his successful movement, Garvey was bound to attract a number of powerful enemies within the ranks of African Americans.

These enemies included black intellectuals, editors, civil rights leaders, political activists, and ministers. His self-appointed efforts to articulate the destiny and feelings of the masses created serious animosity among a number of the educated and affluent black community, particularly the "black church."

The black church remained at the epicenter of the African American community and black activism since the reconstruction period. Although some black preachers endorsed Garvey, numerous black ministers envied the UNIA's effect on their congregations, as black church memberships began to noticeably decline in regions with a strong UNIA presence. To sabotage the impact of the Garvey movement, some preachers resorted to misleading local police and political bosses, as they suggested Garvey was a threat to both black society as well as white society by preaching racial hatred. These exaggerations resulted in arbitrary harassment, threats, and baseless arrests of Garveyites.

In addition to black churches, the National Association for the Advancement of Colored People (NAACP) became another major black institution that sought to discredit and disband Garvey's movement. The NAACP found that its movement, which was grounded upon integration, directly conflicted with Garvey's ideology of black separatism. Garvey countered NAACP criticism by suggesting the organization was not trustworthy because its leadership was nearly all white and he labeled it the "National Association for the Advancement of Certain People." Garvey argued "the NAACP was causing the trouble between the races in the South by giving the impression that the Negro wants intercourse with whites." Garvey suggested, "We stand for a clean cut Negro policy. . . . All we ask is a fair chance in the fields of industry and commerce and we can take care of ourselves" (Garvey and Garvey 1968).

One of the strongest voices against Garvey within the NAACP was the renowned W. E. B. Du Bois, editor of the NAACP's *Crisis*. In fact, Garvey's war of words with Du Bois rose to epic proportions as they frequently exchanged criticism through their periodicals. Garvey stated "DuBois and his allies were planning the extinction of the black race through miscegenation" and were thus "the greatest enemies the black people have in the world" (Garvey and Garvey 1968:). Conversely, Du Bois offered a number of unflattering indictments against Garvey, as he labeled him "a little, fat black man, ugly, but with his intelligent eyes and big head, who was the most dangerous enemy of the Negro race in America and the world . . . who was either a lunatic or a traitor" (quoted from Hill 1986a: 233).

In addition to Du Bois, another noteworthy African American leader who adamantly opposed the separatist Garvey movement was labor leader A. Philip Randolph. In his periodical, *The Messenger*, Randolph labeled Garvey as "the supreme Negro Jamaican Jackass, a monumental monkey," and an "unquestioned fool and ignoramus." Randolph's sentiments about Garvey were complemented by an influential group of African Americans that included Randolph's fellow labor organizer Chandler Owen, along with George W. Harris, Robert

W. Bagnell, William Pickens, Julia P. Coleman, John E. Nail, Robert S. Abbott, and Harry M. Pace. This group of notable African Americans labeled themselves the "Committee of Eight" and poignantly petitioned the U.S. government to arrest Garvey and disband his "dangerous" organization.

When Garvey was eventually arrested on mail fraud charges, the Committee of Eight sent a letter to the attorney general on January 15, 1923, stating:

> Dear Sir:
>
> As chief law enforcement officer of the nation, we wish to call your attention to a heretofore unconsidered menace to harmonious race relationships. There are in our midst certain Negro criminals and potential murderers, both foreign and American born, who are moved and actuated by intense hatred against the white race. These undesirables continually proclaim that all white people are enemies to the Negro.
>
> . . . Its present and moving spirit is one Marcus Garvey, an unscrupulous demagogue, who has ceaselessly and assiduously sought to spread among Negroes distrust and hatred of all white people.
>
> . . . The UNIA is composed chiefly of the most primitive and ignorant element of West Indian and American Negroes. . . . In short, this organization is composed in the main of Negro sharks and Negro fanatics. . . .
>
> For the above reasons we advocate that the Attorney General use his full influence completely to disband and extirpate this vicious movement, and that he vigorously and speedily push the government's case against Marcus Garvey. . . .
>
> We desire the Department of Justice to understand that those who draft this document . . . sound tocsin only because they foresee the gathering storm of race prejudice and sense the imminent menace of this insidious movement, which cancer-like, is gnawing at the very vitals of peace and safety—of civic harmony and interracial concord. (Hill 1986a: 186)

In a *Negro World* article, Garvey called the eight signers "good old darkies" who had "written their names down everlastingly as enemies of their own race." Nonetheless, sentiments like those within the above letter epitomize the tenacious wide-scale campaign that culminated to crush Garvey.

Adding to the initial uproar, the one event that most dramatically strengthened black opposition toward Garvey was his publicized meeting with the Ku Klux Klan (KKK) on June 25, 1922. Before meeting with the KKK, most of the attacks against Garvey from African American leaders were expressed in written smears through newspapers and magazines. After his meeting with the KKK, however, numerous black leaders went to extreme measures in undermining the Garvey movement, which they concluded as perilous to the race. What Garvey planned to accomplish from the meeting with the KKK and what other black leaders perceived of the meeting proved diametrically opposite.

Garvey met with KKK representatives in Atlanta for two major reasons. First, he wanted to disarm the KKK, whose membership was peaking in the early 1920s, and who had initiated episodes of harassment and brutality of UNIA members in the South. UNIA members complained to headquarters that Klansmen suspiciously questioned members on the streets and even appeared at their homes to demand information about the UNIA. Garvey seized the opportunity to explain the similarities in the racial desires of both organizations. Garvey reassured the KKK that his organization opposed miscegenation or any social contact between races. There was significant speculation that the KKK liked Garvey's idea of an African homeland for blacks so much that the KKK agreed to donate money to actualize Garvey's idea of continental separatism. The KKK clearly misinterpreted the scale of Garvey's idea of an African homeland to mean a full-scale black exodus to Africa. At no time of his life did he plan mass or coerced migration. Garvey envisioned select colonization by pioneering people, with the means to establish themselves, and the initiative to help build up a vibrant colony. Ultimately, the UNIA and the KKK failed to reach any significant agreement; furthermore, a Klansman official actually testified against Garvey at his mail fraud trial the following year.

No other act of Garvey's career brought him more criticism and cost him more support than his KKK meeting. Nevertheless, Garvey remained unapologetic, as he continued to view the KKK as a possible enabler to the realization of his ideas. He stated, "Compared with the farce, hypocrisy and lie typical of most whites, we welcome the honesty of purpose of the Klan. They are better friends to my race, for telling us what they are, and what they mean, thereby giving us a chance to stir for ourselves." Additionally, he offered one of his most controversial assessments as he asserted, "Potentially, every white man is a Klansman . . . and there is no use lying about it. The KKK is a help to this movement by increasing the membership, by making the black man think of Africa" (Garvey and Garvey 1968).

Despite Garvey's justification for meeting with KKK officials, mischaracterizations flourished throughout many corners of black America. Numerous black leaders felt Garvey's oversimplification of "all whites" having the same mentality as the KKK as dangerous and would alienate those whites who were sympathetic to their struggle. Most black leaders chose to publicly oppose Garvey, even those black leaders who previously supported some aspects of his organization or philosophy. Carter G. Woodson completely ended his article contributions to the *Negro World* in the aftermath of Garvey's meeting with KKK officials. Similarly, one-time ally William Pickens began to shun Garvey as he refused to attend the August 1922 UNIA convention where he was being honored. Pickens declared, "Wherein I have thought Marcus Garvey to be right I have said so, regardless of the opinions of those opposed to him. Now that I know him to be wrong I say so." In a letter to Garvey, Pickens concluded, "I gather from your

recent plain utterances that you are now endorsing the Ku Klux Klan, or at least conceding the justice of its aim to crush and repress Colored Americans and incidentally other racial and religious groups in the United States" (quoted from Avery 1989: 68). Subsequently, Pickens became a major force in the "Garvey Must Go" campaign. Black integrationists suggested that Garvey encouraged the Ku Klux Klan to brutalize blacks to cause them to want to leave the United States for Africa. Criticism against Garvey grew so significant that opposition even extended outside of the United States to European colonies in Africa.

European powers such as England, Belgium, Italy, and France communicated their concern to the U.S. government about UNIA activities in Africa. They found Garvey's philosophy and encouragement of African independence as unacceptable, particularly in light of the enormous economic benefits they were reaping from their colonized territories in Africa. European powers desperately wanted to deter the effectiveness of the Garvey movement; therefore, numerous UNIA members were harassed and arrested in various countries under petty charges.

Additionally, European powers were clever in promoting their displeasure of Garvey through native Africans such as Blaise Diagne, who was a Senegalese member of the French Chamber of Deputies. Blaise was manipulated by the French government to publicly reject Garvey "in the name of all Africans." Blaise stated:

> We French natives wish to remain French, since France has given us every liberty and since she has unreservedly accepted us upon the same basis as her own European children. None of us aspires to see French Africa delivered exclusively to the Africans as is demanded, though without any authority, by the American Negroes at the head of whom you have placed yourself. (quoted from Rogers 1996: 162)

European powers also outlawed any UNIA activities in their colonies. In fact, European suppression combined with U.S. government efforts to subdue Garvey proved to be the greatest factor leading to the downfall of the Garvey movement.

Indeed, beyond its own desire to persecute Garvey, the U.S. government seized momentum against Garvey, already created by European colonial powers and by the plethora of anti-Garvey rhetoric and activism from black leaders. Accordingly, the U.S. government pursued its own design to persecute Garvey. Via the State Department, the U.S. government directed one of its first official blows by issuing an insidious statement, labeling Garvey "an undesirable and indeed a very dangerous alien, whose aim was to pit all of the Negroes in the world against the white people."

Various conclusions about the government's true motives have emerged from scholars of African American history in recent years. Scholar Derrick McKisick, in a personal interview in 2009, suggested that the U.S. government's true impetus

for pursuing Garvey was grounded in its oversimplified notion that Garvey sought to remove all blacks from the United States to Africa. McKisick contends that U.S. officials were motivated against a black exodus for economic reasons, not moral reasons, as they were unwilling to part with blacks who were a needed homegrown cheap labor force at a time of heavy industrialization. Scholar William Boone, in a personal interview in 2009, asserted that U.S. officials were primarily motivated by racist reasons to stop a black exodus, as U.S. officials worried that a black exodus would radically disrupt the socioeconomic hierarchy grounded in white supremacy, where blacks served a needed purpose by being permanently fixed at the bottom.

Marcus Garvey, in uniform as the president of the Republic of Africa. Garvey founded the Universal Negro Improvement Association (UNIA) to unite African peoples worldwide and encourage African Americans to return to Africa. (Library of Congress)

Whatever their primary motivations, federal authorities arrested Marcus Garvey and three of his associates on January 12, 1922, and indicted them on 12 counts of fraudulent use of the mail to sell Black Star Line stocks. The three Black Star Line officials charged with Garvey were Eli Garcia, George Tobias, and Orlando Thompson. The actual trial began over a year later on May 18, 1923.

The prosecution's opening statement included "the entire scheme of uplift was used to persuade Negroes for the most part to buy shares of stock in the Black Star Line . . . when the defendants well knew . . . that said shares were not and in all human probability never could be worth $5 each or any other sum of money." Even before the trial began, there was a heated exchange between UNIA lawyers and the federal judge presiding over the case, Julian Mack. The defense attorneys insisted that Judge Mack remove himself from the trial due to his membership in one of the major anti-Garvey organizations, the NAACP. Judge Mack admitted his association with the NAACP, yet refused to step down, claiming he would not be biased.

Two days into the trial, Garvey dismissed his lawyer and decided to defend himself. Although Garvey knew the move could be detrimental to the outcome of his case, he stood firm on removing UNIA Counsel General William Matthews.

Garvey became outraged at Matthews's insistence to accept a plea bargain for a reduced charge. Garvey's co-defendants, however, maintained their original legal representation.

At times, Garvey effectively provided solid defense and cross-examination, however, courtroom onlookers characterized most of Garvey's efforts as blundering, legally incompetent, and even comical. Garvey's overt mistakes in defending himself were compounded by persistent condescending comments from Judge Mack. From the beginning of the court proceedings, Judge Mack's treatment of Garvey caused many interested in the trial to believe that Garvey was doomed.

The federal government provided a number of ex-Garvey associates and employees to testify for the prosecution. The judge allowed their testimonies to go far beyond the mail-fraud charges, as witnesses testified to unrelated matters such as Garvey's occasional horse race betting when visiting Jamaica in 1921. The prosecution also relied on the contrived accounts of former UNIA officer Adrian Johnson, who claimed officials of the international UNIA office were secretly paid large sums. Years later Johnson sued the UNIA, claiming he had hardly been paid at all for his year of services; thus showing that his previous testimony was erroneous.

During his defense, Garvey stated that his organizations were "more intent on the ultimate uplifting and salvation that was promised to the Negro race of America than in the paltry profits that might be realized from the stock investment." Additionally, one of the lawyers for Garvey's co-defendants reaffirmed Garvey's stance by lamenting that "If every Negro could have put every dime, every penny into the sea, and if he might get in exchange the knowledge that he was somebody, that he meant something in the world, he would gladly do it. . . . The Black Star Line was a loss in money but it was a gain in soul" (Ottley 1967: 220).

The government's official accusations, that the Black Star Line used the U.S. mail to defraud the public, seemed limited in light of the inconsistent evidence provided, especially the contradictory testimonies. To build its case, the prosecution provided various individuals who contended to be victimized by Garvey's organization. Only one of the prosecution's witnesses, however, purchased stock through the mail. There were even a number of the prosecution's witnesses who had not purchased stock at all. Nevertheless, the federal government consistently placed people on the witness stand who espoused negativity about Garvey or the Black Star Line, despite their lack of actual firsthand evidence to support the government's charges.

The key witness for the prosecution, Benn Dancy, unveiled an envelope mailed to him from the Black Star Line. Although the envelope was authentically addressed from the Black Star Line, Dancy could not provide any contents of the envelope. Furthermore, there were notable accounts from onlookers that Dancy's testimony was incoherent and inconsistent. The blatant ineptitude and

contradictions within Dancy's testimony led to the exoneration of Garcia, Tobias, and Thompson. Yet Dancy's testimony, with its monumental flaws, was the key element in convicting Garvey. After being found guilty by a jury, Garvey received a five-year prison sentence on June 21, 1923. How Garvey was found guilty, yet his co-defendants found innocent of the same charges, revealed a level of prejudice and complicity between the federal government and the judicial system. In the face of witness coercion and shaky evidence, railroading Garvey into a conviction and isolation seemed paramount to all involved in the trial, including judge, jury, and prosecution. Historian Lawrence Levine contends that Garvey was guilty of nothing more than "poor management, inexperience, and bad judgment in choosing some of his associates" (1982: 134), as only Garvey of the four defendants was found guilty and sentenced to five years in prison.

Despite Garvey's courtroom mistakes, he, in addition to the lawyers for his co-defendants, masterfully exposed the prosecution's attempts in coaching witnesses to fabricate testimony, which was a definite basis for a mistrial. For example, prosecution witness Schuyler Cargill testified to being employed as a mail clerk under the supervision of "Mr. Prentice" in the Black Star Line offices in 1919. Soon after Garvey proved that Prentice was not working for the Black Star Line at the time that Cargill testified about, Cargill readily admitted that the prosecutor fed him the dates to further fabricate the case against Garvey. Despite this proven false testimony and obvious witness coercion, no mistrial was declared and the trial of Garvey and his co-defendants was allowed to continue.

In cases identical to Garvey's, those convicted of mail fraud were typically sentenced to probation or up to one year of incarceration. The fact that Garvey was given the maximum five-year sentence, based on such limited evidence and a questionable trial, exhibited the federal government's and judicial system's obsession with ostracizing him for as long as possible, in hopes of destroying his movement. Directly following sentencing, Garvey was held without bail for three months at Tombs Prison in New York, although legally Garvey should have been granted bail more expediently while his case was being appealed.

The trial of Marcus Garvey fully exhibited the power and rage of the forces that wanted to crush Garvey altogether. Beyond the government's insincere concerns that Garvey victimized members via mail fraud, his arrest and conviction unveiled the true intentions of his opposition, to discredit his Pan-Africanist philosophy and dismantle his movement, as the evidence against Garvey proved to be scant. In fact, while incarcerated in New York in June 1923, Garvey accurately concluded, "I am not here because I committed any crime against society or defrauded anyone, but because I have led the way to Africa's redemption" (Garvey and Garvey 1968). Garvey was released on bail in September 1923 on appeal. Garvey's appeal was officially denied on February 2, 1925, as he was imprisoned once again, this time in Atlanta.

The impact of Garvey's arrest and conviction proved catastrophic, as his movement never recovered, signifying victory for the U.S. government and all of Garvey's enemies. Specifically, his legal proceedings resulted in the loss of valuable human resources. Governmental infiltration and successful efforts to sabotage Garvey's movement heightened his paranoia about those close to him. Accordingly, Garvey accused a number of noteworthy UNIA officials and members of being self-serving or disloyal. The government's attempts to create in-fighting worked masterfully and resulted in the resignations of notable UNIA officials. Those who resigned included Garvey's original trial lawyer, his three trial co-defendants, and their lawyers; all of whom felt unappreciated or attacked by Garvey. Unfortunately for the Garvey movement, these individuals also served as auditor general and the bulk of the UNIA legal staff. Garvey's overall constituency also dramatically declined throughout the United States as a result of his conviction.

Furthermore, after Garvey's conviction, the almost bankrupt UNIA soon became plagued with pending lawsuits totaling almost a quarter million dollars. The economic standing of the UNIA grew so dismal that the final ship in Garvey's fleet was sold in an attempt to cover some of the debt accrued. Compounding the troubles, the UNIA lost a number of major buildings such as liberty halls, office buildings, and publishing plants. Thereafter, the UNIA only functioned on a minimal basis as a shadow of its previous glory.

Financial disarray and severe membership erosion were directly related, and both could be attributed to the effectiveness of government repression. One of the severest blows to the organization and rank-and-file was the U.S. government's confiscation of records, which were never returned. Numerous members loaned significant amounts of money to the UNIA, but proof of their contributions was lost with the seizure of records. The loss of records led many members to grow so angry at the UNIA that many left the movement altogether. The shattering financial and record-seizing situation deterred potential investors as well, which prevented possible new income for the UNIA.

Garvey and others credit federal agents planted within the UNIA with spreading rumors that the Black Star Line and UNIA officials were embezzling funds. Many Garveyites, unaware of the government's efforts in sabotage, skeptically began to question UNIA leadership during the chaotic times. All of these combined effective forms of government infiltration and repression led to severe recruitment difficulty, not only in the United States, but also in the Caribbean and Africa. Most of the Garveyites who chose to leave the movement during this tumultuous time still maintained strong convictions in black nationalism as a philosophy or a goal, yet were moved to leave the UNIA over perceived poor management. Ultimately, due to legalities and harassment, Garvey was forced to base his headquarters outside the United States. Operating out of London and Jamaica, away from his powerbase in the United States, proved to be one of the final factors that led to the death of Garvey's movement.

Race and the Federal Bureau of Investigation

James Wormley Jones was one of the first five African American agents hired by the Federal Bureau of Investigation (FBI), and he was involved in the case that led to Marcus Garvey being jailed for mail fraud. He was born on September 22, 1884, at Fort Monroe, Virginia, the son of John and Sal Jones. His family moved to Cambridge, Massachusetts, when he was young, but they returned to Virginia, with John Jones becoming a light housekeeper at Chesapeake, with James Wormley Jones attending Norfolk Mission College and the Virginia Union University.

In January 1905, Jones started working with the Washington Metropolitan Police Department, starting as a footman, and then rode a horse, and finally a motorcycle. He was later promoted to being a detective, and in World War I he served in the U.S. army, gaining a commission, and then serving in France.

Returning to the police when he came back to the United States, Jones was appointed to the FBI in 1919. He was responsible for infiltrating the Universal Negro Improvement Association, and there he sought evidence on Garvey. Jones resigned from the FBI in 1923 and went to live in Dormont, Pennsylvania, where he died on December 11, 1958.

Garvey became even more disheartened when he learned that his plan of setting up a colony in Liberia was thwarted. His plan of actualizing African independence and the "Back to Africa" movement began with acquisition of land in Liberia. In December 1923, before Garvey was released on bail, UNIA delegations traveled to Liberia to negotiate land acquisition toward the ultimate goal of black settlement. Garvey was adamant about establishing a homeland for blacks of the diaspora to visit or resettle. Initial UNIA negotiations with Liberian officials proved to be very promising. After Garvey's incarceration, however, Liberian officials completely changed their previous agreement. Then, when UNIA officials arrived in late 1924 with money and supplies, Liberian forces quickly apprehended them and sent them back to America. Intense coercion from European colonial powers such as England, France, Italy, and Belgium, in addition to the United States, became the catalyst behind Liberia's retraction of its initial goodwill toward Garvey.

The Liberian government punctuated their sentiments by declaring that they would immediately apprehend Garvey and surrender him to British authorities if he attempted to personally visit to resolve issues. Garvey, however, was never afforded the opportunity to personally attempt renegotiations with Liberia because he lost his court appeal. Thereafter, he was promptly apprehended by U.S. authorities on February 8, 1925, and taken to Atlanta Federal Penitentiary. During his resulting two and a half years of imprisonment, the UNIA almost completely dissipated.

In a February 10, 1925 letter, one of Garvey's many messages that he offered while incarcerated, the fallen leader stated "If I die in Atlanta my work shall then only begin" (Hill 1986b: 98). Despite Garvey's justifiable paranoia, he did not die

while imprisoned. Ostracism of a leader, sometimes an outcome as severe to a movement as death of a leader, would be Garvey's fate, as his imprisonment transitioned into banishment from the United States. Two years before Garvey's sentence was scheduled to end, President Calvin Coolidge commuted Garvey's prison term on November 18, 1927, and within the next month he was deported as an undesirable alien. As his ship departed a New Orleans port, Garvey told onlookers, "Be not dismayed! . . . Africa's sun is steadily and surely rising, and soon shall shed its rays around the world. I live and shall die for Africa redeemed. Steady yourselves and go forward!" (Garvey and Garvey 1968).

The trial and conviction of Marcus Garvey for defrauding his followers destroyed his worldwide movement of black independence from white oppression. Garvey's conviction resulted more from his provocative concepts than from the evidence, which was slight. He organized a mass movement of people of African descent larger than any seen before or since. His millions of followers in the United States, the West Indies, Latin America, and Africa were considered a significant threat by the European powers that controlled Africa. Garvey not only pioneered the idea that "Black is beautiful," but he also set Africa's liberation from white domination as his goal. Unwilling to allow significant numbers of African Americans to exit the United States, and acquiescing to the anti-Garvey momentum, the federal government railroaded Garvey into prison on trumped-up mail fraud charges in 1925 and eventually exiled him out of the United States. By 1935, organized suppression, infiltration, and ostracism effectively led to the dismantling of the UNIA in every country.

References and Further Reading

Avery, Sheldon. *Up from Washington: William Pickens and the Negro Struggle for Equality*. Newark: University of Delaware Press, 1989.

Garvey, Amy Jaques, ed. *Garvey and Garveyism*. Kingston, Jamaica: United Printers, 1963.

Garvey, Marcus, and Amy Jaques Garvey, eds. *The American Negro: His History and Literature: Philosophy and Opinions*. New York: Arno Press, 1968.

Hill, Robert A., ed. *The Marcus Garvey and Universal Negro Improvement Association Papers:* Vol. 5, September 1922–August 1924. Berkeley: University of California Press, 1986.

Hill, Robert A., ed. *The Marcus Garvey and Universal Negro Improvement Association Papers:* Vol. 6, September 1924–December 1927. Berkeley: University of California Press, 1986.

Levine, Lawrence. "Marcus Garvey and the Politics of Radicalization." In *Black Leaders of the Twentieth Century* (pp. 105–8). Edited by John Franklin Hope and August Meier. Urbana: University of Illinois Press, 1982.

Maglangbayan, Shawna. *Garvey, Lumumba, Malcolm: National Separatists*. Chicago: Third World Press, 1972.

Ottley, Roi and William J. Weatherby. *The Negro in New York: An Informal Social History*. New York: New York Public Library, 1967.

Rogers, J. A. *World's Great Men of Color*, Vol. 2. New York: Touchstone, 1996.

Vincent, Theodore. *Black Power and the Garvey Movement*. New York: Ramparts Press, 1972.

CON

Marcus Garvey was one of the leading proponents of African American racial consciousness during the 1920s. He called for equal rights for African Americans and an end to practices such as lynching. Garvey also encouraged African Americans to study and value their own culture and self-worth. He called for a return to Africa, in an effort to make that continent free of colonial domination and to establish true democracies.

Garvey caused many white Americans great concern. They saw him as a radical agitator who threatened the status quo. His rejection of accepted social practices made Garvey an object of suspicion. The World War I and postwar era was a time of great social upheaval, and some Americans saw Garvey as a part of that threat. Other radicals were persecuted with extralegal means. Garvey, however, received the benefit of due process. Although some questioned whether the charges against him were valid, Garvey was allowed to defend himself and to appeal his conviction. He was sentenced to a normal prison term, which was later commuted. Garvey was deported, which was, and is, the ordinary practice for an alien convicted of a felony.

Historical Introduction to the Time

During the late 19th and early 20th centuries, American society underwent a series of upheavals. Much of the impetus was the growing importance of radical movements, especially those that championed labor and other oppressed groups. Movements such as anarchism, socialism, and Marxism sought to overturn the traditional social order. They called for changes such as government intervention and control over production and working conditions. The reformers also wanted workers to share in the profits of their labor. Labor unions grew in importance, resulting in many strikes. Many of these strikes ended in bloodshed. The radical organizations included groups such as the Industrial Workers of the World (IWW), known as the "Wobblies" to their opponents. A number of the IWW's leaders were jailed or executed under questionable circumstances.

Many of the more radical leaders were recent immigrants, causing further concern among the upper and middle classes. The United States experienced great population growth during this period, mostly through immigration. Most of these immigrants were not from Western Europe and other traditional sources of U.S. immigrants. Many came from southern and eastern Europe, with ideas and philosophies foreign to those traditional in U.S. society. Suspicion of these recent immigrants turned to alarm when they espoused anarchism and socialism and called for changes in the U.S. system.

Government and social leaders seemed to have their suspicions confirmed during World War I. The stresses of war caused social change throughout Europe. In Russia in particular, radical change took place. The Bolsheviks seized control of the country. They executed the tsar and his family and made peace with the Germans who were still fighting the Americans. Other European countries, such as Germany and Austria-Hungary, experienced revolt by the working classes and the overthrow of established governments.

Many in the United States feared radicals planned the same kind of changes in this country. During the war, legislation was passed that limited free speech and activities. Private organizations monitored the statements of individuals and reported violations to law enforcement. A number of Americans who opposed the war, such as Eugene Debs, were arrested and sent to prison for extended terms. Such harsh actions seemed justified by acts of terrorism conducted after the war ended in November 1918. On June 2, 1919, a number of bombs went off in eight cities, including Washington, D.C. Attorney General A. Mitchell Palmer was targeted by one, and his home was partially destroyed. On September 16, 1920, a bomb estimated to have 100 pounds of dynamite and 500 pounds of steel fragments was set off in front of J. P. Morgan's bank on Wall Street. Thirty-eight people were killed and 400 others were injured.

The federal government responded with raids against anarchists, socialists, and Marxists, and those suspected of being sympathetic to those movements. Hundreds were arrested and often given cursory trials. Many were deported, such as Emma Goldman. Public opinion was very much opposed to the radical movements.

Conditions for African Americans

African Americans were not generally considered likely members of the radical movements, but most white Americans were concerned about the development of new attitudes among African Americans. After all, African Americans were freed from slavery at the end of the Civil War, but their condition had generally worsened. Many blacks were kept in poverty and deprived of their civil rights. "Jim Crow" laws enforced the subservient position of African Americans in relation to whites. Although many progressives campaigned for greater equality,

by 1915, a conservative reaction was taking place in American society. That year, the Ku Klux Klan was reborn in Georgia. It was inspired by the release of the film *The Birth of a Nation* and by the lynching of Leo Frank, a Jew convicted of raping and murdering a girl who worked at his factory. The Klan's program of anti-Catholic, antiforeigner, anti-Jewish, and anti-African American activities met with considerable approval in much of the United States.

For many African Americans, the failure of state and federal governments to protect them was embodied by the practice of lynching. In 1915, 56 blacks were murdered by crowds of whites. Lynching of African Americans reached a high of 76 in 1919 before beginning a gradual decline. The crimes of the victims of lynching ranged from being accused of rape or murder of white women to simply being disrespectful. Lynching was accompanied by race riots during the time period. A representative disturbance was the 1917 race riot in East Saint Louis. Beginning on July 2, 1917, the root causes were economic as well as racial. With the outbreak of World War I, U.S. industries began a major expansion. Workers were in short supply, which encouraged many blacks from the South to migrate to northern industrial cities. East Saint Louis was a center for manufacturing and had always had a small African American population. By 1917, the number of African Americans living there had increased many times over. Employers signaled they could hire blacks for less money and that they would do jobs white workers would not. To ease tensions, city authorities placed an embargo on the sale of guns to African Americans. Those entering the city were searched. Some light-skinned blacks were able to pass themselves off as white and carried weapons through police lines. Rumors flew that the African Americans were planning to massacre whites on July 4. On July 2, mobs of whites formed and attacked blacks they encountered on the street, before attacking African American neighborhoods. Firing broke out from both sides. By the time the Illinois National Guard put an end to the fighting, the official death toll was 39 African Americans and 9 whites. The real number of deaths was much higher. Congressional hearings into the riots resulted in condemnations, but no real changes. Many Americans were sympathetic to a cartoon printed shortly afterward. It showed an African American woman and her children appealing to President Woodrow Wilson for democracy in America, even while Wilson holds his promise to wage war to make the world safe for democracy.

Marcus Garvey

Marcus Garvey was born in Jamaica in 1887. By trade, he was a printer, but he became interested in labor and racial issues after being blacklisted following his participation in a strike. Between 1908 and 1912, he traveled throughout the Caribbean and observed the conditions under which blacks worked and lived. From 1912 to 1914, Garvey lived in England, where he met a number of African

nationalists, studied philosophy, and read widely. He became convinced that blacks could only improve their condition by uniting to create economic power. He also believed blacks should become more aware of their racial heritage and adopt a stance of black nationalism. When Garvey returned to Jamaica in July 1914, he quickly founded the Universal Negro Improvement Association (UNIA) and the African Communities League (ACL). The UNIA was intended to create economic opportunities for blacks, while the ACL worked to end European colonialism in Africa. The slogan for UNIA was "One God, One Aim, One Destiny."

Garvey corresponded with Booker T. Washington in 1915 and learned about conditions in the United States. He arrived in America on March 23, 1916, hoping to raise money through a speaking tour for a school similar to Washington's Tuskegee Institute. Although Washington had died in 1915, Garvey visited Tuskegee and became acquainted with many African American leaders. He perceived what he believed was a lack of leadership among African Americans and soon embarked on a 38-state speaking tour to champion his philosophy.

In May 1917, Garvey and 13 others formed a division of the UNIA in New York. The East Saint Louis riot followed only two months later. Garvey was prompted to speak out against the discrimination and ill-treatment that caused the riot. Under the UNIA's auspices, Garvey began to work for the improvement of African Americans' conditions. He started to publish the widely read *Negro World* newspaper on August 17, 1918. By June 1919, UNIA had 2 million members.

In November 1918, Garvey came to the attention of federal officials when the British War Office asked for information about him and his connection with African nationalists in Great Britain. Emmett Scott, a black official in the War Department who had met Garvey, called him to Washington to discuss Garvey's work. He warned Garvey that comments in the *Negro World* about lack of unity of purpose among the racial groups were provocative. Even so, Scott tended to dismiss Garvey as a dangerous agitator.

Other federal agents were assigned to attend UNIA meetings and report back on what was said. Major Walter H. Loving, a retired black officer, had greater concerns than Scott, based on what his agents reported back. In January 1919, Loving sent word to his superiors that he believed that Garvey's speeches and editorials would have been considered seditious if he had spoken them during the war. Garvey had written that the black race would side with whichever side offered them freedom and liberty. With the peace, however, wartime restrictions began to be relaxed, although the federal government remained vigilant about agitators.

Some members of the Justice Department wanted to prosecute Garvey by early 1919, but the assistant attorney general in charge of prosecutions for sedition believed Garvey's statements were permissible attempts to influence government policy in a democracy. Most of the Justice Department's efforts during 1919 were directed toward monitoring foreign-born socialists, anarchists, and Marxists. J. Edgar Hoover was head of the General Intelligence Division, and

he used his manpower to monitor those groups. It was only after the "Red Scare" began to decline that Hoover paid additional attention to African Americans like Garvey. The first five African American agents of what became the Federal Bureau of Investigation were hired to keep tabs on Garvey's messages to his followers.

Race riots in Washington and Chicago in July 1919 prompted the army's Military Intelligence Division and the Federal Bureau of Investigation to authorize the recruitment of "reliable Negroes" to attend meetings of African American organizations. The army in particular developed a sophisticated analysis of the mood of African Americans in 1919. Military intelligence determined that the regimentation of society during World War I had encouraged the development of a black identity. African Americans became more aware of their combined strength and that they were expected to sacrifice while receiving little in return. They were upset over lynchings in particular, along with Jim Crow laws and discrimination. The army supported the development of interracial organizations like the Urban League that could improve relations between the races.

Officials at the Justice Department, including Hoover, were concerned that black nationalism could merge with other forms of radicalism. Hoover favored indictments of African American leaders, including Garvey, but federal attorneys were opposed since they did not believe they could obtain convictions for the proposed crimes. The presumed danger of Garvey was rated as low.

During 1919, Garvey put his program of economic empowerment into practice. One of the most visible was the founding of the Black Star Line in June 1919. Based on the name of the famous White Star Line, Garvey intended this business to be an African American–owned and –operated shipping line. The company was intended to carry merchandise between the United States, the Caribbean, and Africa. Donations from UNIA members enabled Garvey to purchase a ship shortly afterward. With much publicity, the SS *Yarmouth* was purchased and renamed the SS *Frederick Douglas* on September 14, 1919. Other ships followed. The shipping company, however, was never well run, and some of the ships that were purchased were found to be unsuitable.

Garvey also founded the Negro Factories Corporation as a part of the UNIA. The corporation was intended to produce all kinds of marketable commodities in factories all over the United States. The factories would be owned and operated by African Americans and employ African Americans at wages equivalent to those received by white workers. Garvey hoped the business would eventually include a chain of grocery stores, restaurants, and other businesses where blacks could receive service sometimes refused them by white owners.

Garvey's efforts were enormously successful. It has been estimated that more African Americans belonged and participated in the UNIA than in the civil rights movement of the 1950s and 1960s. By August 1920, the UNIA had 4 million members. That month, the first international convention of the UNIA

was held in Madison Square Garden. Over 25,000 people heard Garvey speak about his dreams on August 1. Garvey used other public events to raise race consciousness and pride among African Americans. From his headquarters in Harlem, Garvey organized massive parades and demonstrations of the UNIA. Wearing colorful uniforms, Garvey and other leaders of the UNIA were conspicuous at these events.

Garvey also included a "back-to-Africa" element in his philosophy. He believed that Africa should be freed from colonial rule and that its infrastructure should be developed. He developed his "Liberia Plan" to build factories, schools, and railroads in that country. By helping Liberia to develop economically, it could become a center to inspire democracy in the rest of Africa. At the international convention in 1920, the delegates adopted a constitution, a Declaration of the Rights of the Negro Peoples of the World, and a national flag for the country they hoped to create in Africa. Garvey was elected provisional president of this African country.

Garvey's Arrest and Trial

Garvey's success in organizing African Americans obviously concerned many whites. Although some dismissed him as a buffoon or showman or conman, by 1920, Garvey wielded a great deal of power. His plans for the economic development of black purchasing power threatened white businesses, while his plans to keep the races separate was unacceptable to many African American and white liberals. Garvey had many enemies in the African American community. The most famous was W. E. B. Du Bois. Du Bois believed the future of African Americans lay in integration into mainstream American society. The dispute soon degenerated into personal attacks. Garvey believed Du Bois disliked him because his skin was a darker shade and he was from the Caribbean. Du Bois called Garvey stupid and intimated that he was profiting unjustly from the UNIA.

Federal agencies continued to investigate Garvey and whether he was guilty of threats to society. Public opinion had turned against the limits on freedom of speech dating from World War I, so federal agents could not simply accuse him of seditious speech. Investigations into the finances of the UNIA and Garvey's tax records found nothing. It was only in 1922 that a possible crime was uncovered, thanks to the U.S. Post Office. Garvey was arrested on January 12, along with three other officers of the Black Star Line. They were indicted on February 16 for mail fraud.

Garvey and his associates were accused of soliciting funds for the line by claiming to own a ship that had not been purchased. Fliers with a picture of the SS *Orion* were sent through the mail to members of the UNIA, with offers to purchase stock in the Black Star Line. The ship in the picture had the name "Phyllis Wheatley" on it. The Black Star Line planned to purchase the ship, but when the

flier was sent out, it had not yet been bought. The flier indicated that it was already owned by the line, leading to the charge of fraud.

The investigation leading up to the trial brought charges of wrongdoing from many of Garvey's enemies. Some were former associates in the UNIA, such as Gwendolyn Campbell, Garvey's personal stenographer. She testified at the trial that Garvey and his wife dipped into UNIA funds for their own use. Other members accused Garvey of being a swindler. All of these charges became public when the trial opened on May 21, 1923.

Garvey did not help himself in the months before the trial. When released on bail, he failed to deny that any wrongdoing occurred. Instead, Garvey blamed the others arrested with him. By doing so, he

Marcus Garvey leaves court handcuffed to a deputy, New York City, February 6, 1925. Garvey was sentenced to five years in Atlanta Penitentiary for mail fraud. (Bettmann/Corbis)

implicitly acknowledged that a crime had taken place without him being aware of it. Testimony during the trial, however, portrayed Garvey as an administrator who dominated his subordinates and did not allow anything to happen without his approval. This weakened his defense of not knowing.

Garvey also failed to enlist the aid of his followers. Whenever he spoke, he displayed such confidence that he would not be found guilty that the initial anger over his indictment dissipated. No demonstrations or efforts to get the federal government to dismiss the charge took place. Instead, Garvey enlisted the help of influential African Americans to work behind the scenes to try to prevent any trial from taking place. He had such confidence in his ability to beat the charges that he authorized the continued sale of stock in the Black Star Line.

In his opening statement, prosecutor Maxwell Mattuck concentrated on the legal issues of the case, not any political motivations for the government's charges. He charged that Garvey and his associates knowingly sold stock without any hope for the shipping line's success and without any intention of serving the African American people as a whole. The witnesses testified about Garvey's methods of administering the UNIA and other businesses. They also produced evidence that Garvey enjoyed a lifestyle beyond that of most black men, such as fancy clothing and traveling in first-class accommodations. The testimony also

revealed that UNIA and Black Star Lines were approaching bankruptcy, either through mismanagement or because of theft.

The main evidence produced by the government was an empty envelope from the Black Star Line addressed to an African American named Benny Dancy who lived in Harlem. The government claimed it contained the flier with the fraudulent picture of the *Phyllis Wheatley*, with an offer to purchase stock. The evidence was weak, but Garvey was unable to shake the idea that the envelope had included a fraudulent stock offer.

Garvey decided after the first day of court to argue his own case. After firing his own lawyer, he was aided by the lawyers of his other defendants. They helped Garvey with legal strategy and went over the points to make in each day's testimony. Judge Julian W. Mack also helped Garvey by giving him careful instructions about his rights and privileges in court. The trial followed the normal procedures. Garvey's cause was undoubtedly hampered by his behavior in court. When confronted by unfriendly witnesses, he tended to be critical and aggressive and made personal attacks. His caustic questions also led some witnesses to give unfavorable opinions about his actions and honesty.

Garvey's final summation may also have made the jury question his motives. Although he stated that he had founded the Black Star Line to help his race, some jurors saw him as using race as a cover for shady business practices. The impression may have led to his conviction.

Judge Mack did his best to run a fair court. He even praised Garvey's summation of the facts at the end to the jury. In his instructions to the jury, Mack told them to find Garvey innocent if they believed he was motivated by racial patriotism or the desire to make a profit. They were only to find Garvey guilty if they believed he had been motivated by greed. In the end, Garvey was convicted of mail fraud while his co-defendants were found not guilty. Immediately after the verdict was announced, Garvey responded with "Damn the Jews." This anti-Semitic remark may have led Mack, who was Jewish, to hand down the most severe punishment possible. He sentenced Garvey to five years in prison, a $1,000 fine, court costs, and denial of bail.

Despite the concerns of governmental officials, no violence resulted from Garvey's conviction. Garvey's followers were upset, but remained law-abiding citizens. Garvey blamed his African American enemies, the colonial powers such as France, the Jews, and the whites. He did not, however, blame the federal government for an unjust conviction. Garvey was eventually granted bail on September 10, 1923, while his case was on appeal. Garvey spent the next few months traveling around the country, speaking to African American crowds and raising money for his appeal. He even led the fourth annual international convention of the UNIA in Harlem in August 1924.

Garvey hired the politically influential law firm of Battle, Vandiver, Levy and Van Tine to prepare an appeal to the U.S. Court of Appeals. The grounds

were Judge Mack's technical instructions to the jury. Specifically, they argued that the federal government had not proved its case that Benny Dancy had received an offer from the Black Star Line to purchase stock. The Appeals Court turned Garvey's appeal on February 2, 1925. His attorneys immediately appealed to the U.S. Supreme Court. The higher court refused to hear the case on March 23.

Garvey was arrested on his way back to turn himself in to authorities, on February 5, 1925. He was sent to the federal prison in Atlanta on February 7 and began serving his sentence. Garvey's wife and followers waged a petition campaign to have him pardoned. By 1927, even many of Garvey's opponents had begun to question his conviction. African American leaders threw their support behind a pardon for Garvey. Liberals believed the conviction was unfair. Even segregationists supported Garvey's release; they shared some of his ideas about the separation of the races. On November 18, 1927, President Calvin Coolidge commuted Garvey's sentence to the time already served. Upon release, however, Garvey was subject to immediate deportation. Federal law required that an alien convicted of a felony would be deported after serving his sentence. Garvey had applied for U.S. citizenship, but had never been granted it. On December 2, 1927, Garvey was placed aboard a ship on its way to Panama and then Jamaica. He never again set foot in the United States.

The UNIA continued to function after Garvey's deportation. Many African American leaders of the coming decades were influenced by Garvey and his movement. Elijah Muhammad, founder of the Nation of Islam, was a follower of Garvey. Malcolm X's parents were members of UNIA and passed along much of Garvey's philosophy to him. Others have paid tribute to Garvey, including the Rastafarians, who regard him as a prophet.

Conclusion

Marcus Garvey's trial was fundamentally fair. He was not railroaded into prison, even if doubts exist about the crime for which he was convicted.

In the context of developments between 1910 to 1920, many Americans were concerned about social change and threats by radical groups to the existing order of things. Labor unions, socialists, and anarchists threatened the capitalist system under which the U.S. economy operated. Many leaders of these radical groups were foreign-born, leading to suspicions of those not born in this country. Terrorist acts linked to these groups helped increase the concern of authorities.

The events of World War I also caused more emphasis to be placed upon the protection of the status quo. The Bolsheviks overthrew the government in Russia and took it out of the war against Germany. They also executed the tzar and began a process to confiscate the means of production in that country. Bolshevik and socialist movements in other European countries threatened the

existing governments, causing fears of a worldwide revolution. The federal government responded by limiting civil rights in the United States and by prosecuting those suspected of wanting to change the system.

African Americans during this time migrated in increasing numbers to the industrial centers of the North. They were looking for economic opportunity, but the changing relationships between the races made many whites uncomfortable. Blacks who wanted to change the discrimination and injustices under which they had suffered were viewed with suspicion by many whites. Lynching was only the most visible injustice suffered by African Americans. Racial riots, such as the one that occurred in East Saint Louis, made both blacks and whites fearful of what could happen.

Marcus Garvey entered into this situation. A foreigner, he preached a philosophy of black nationalism, economic opportunity, and return to Africa for African Americans. His message was extremely popular, attracting up to 4 million followers only four years after he had arrived in this country. Garvey inspired many African Americans to study their heritage and to take pride in their race. The Harlem Renaissance of the 1920s flows directly from his preaching. Garvey himself had many faults. He was showy and domineering. He made many enemies among African Americans who wanted to integrate blacks into the mainstream of U.S. society. His separation of the races actually won considerable sympathy among white segregationists. Government agencies were suspicious of his success in organizing African Americans into a movement that could upset the status quo.

Garvey and the UNIA were under surveillance by government agents for years. This practice, however, was not unique to Garvey. Federal officials investigated Garvey for possible crimes under which he could be prosecuted, but this was again not unique. After the end of the war and the decline of the threat from radical groups, agencies were less able to persecute those who spoke out against government policies. Requests by Justice Department officials to have Garvey arrested were denied because prosecutors realized they had no case against him.

The crime for which Garvey was eventually tried was the fraudulent offering of stock through the mail. The charge was based on a technicality of whether the Black Star Line actually owned the ship its literature implied was an asset of the company. The government's case was weak, but Garvey was not denied any of his civil rights. He failed to use all the resources at this disposal to defend himself, such as making appeals to his followers to campaign to have the charges dropped. Garvey also harmed his case by deciding to represent himself. The decision, however, was his own. He did receive legal advice from the lawyers of other defendants and with considerable prompting from the judge at his trial. Judge Mack's praise of Garvey's summation and his instructions to the jury that allowed racial patriotism as a reason for acquittal were indications of his efforts to give Garvey a fair trial.

The Atlanta Penitentiary and Its Famous Inmates

After being found guilty of mail fraud, Marcus Garvey was sent to the U.S. Penitentiary in Atlanta, Georgia, and was there from February 1925 until November 1927. The prison took three years to build, and it covers 300 acres, with the capacity of 1,200 prisoners. Opening in 1902, Marcus Garvey is one of many famous inmates, and the prison had already become known around the world in November 1920. In the U.S. presidential elections of that year, Eugene V. Debs of the Socialist Party managed to get 919,799 votes while a prisoner in the Atlanta Penitentiary for having urged resistance to conscription in World War I.

Before Debs, Carlo Ponzi, the inventor of the famous Ponzi scheme, was in the prison from 1911 until 1913; and Julian Hawthorne, son of the famous writer Nathaniel Hawthorne, spent a year in the prison (1913–1914), like Garvey, on charges of mail fraud, after selling shares in a nonexistent silver mine. In May 1932 Al Capone was also sent to the prison before being transferred to another jail, and then to Alcatraz.

Garvey also was allowed to appeal his case. The appeals process took a year and a half, during which Garvey was allowed to speak and travel publicly. Garvey himself did not blame the government for his conviction, only its agents. He also blamed his white and black enemies, as well as international forces. The law firm that handled Garvey's appeal was well known and politically influential. The appeal itself was apparently handled without prejudice by the Appeals Court. The U.S. Supreme Court declined to hear the appeal, which is more common than not.

Garvey's sentence was the maximum allowable, but it may have been provoked by his anti-Semitic remarks at the end of his trial. When his sentence was commuted after only two years, federal law was strictly followed in his deportation.

Although many may still question whether Garvey actually committed, or intended to commit mail fraud, he received a fair trial that followed the letter of the law of the day. His treatment by the legal system was completely in accord with the established rules and practices. This treatment is all the more amazing in view of the climate of the times.

References and Further Reading

Burkett, Randall K. *Garveyism as a Religious Movement: The Institutionalization of a Black Civil Religion.* Metuchen, NJ: Scarecrow Press, 1978.

Cronon, Edmund David, comp. *Marcus Garvey.* Englewood Cliffs, NJ: Prentice-Hall, 1973.

Du Bois, W. E. B., et al. *African American Political Thought, 1890–1930: Washington, Du Bois, Garvey, and Randolph.* Armonk, NY: Sharpe, 1995.

Fax, Elton C. *Garvey: The Story of a Pioneer Black Nationalist*. New York: Dodd, Mead, 1972.

Hill, Robert A., and Barbara Blair. *Marcus Garvey, Life and Lessons: A Centennial Companion to the Marcus Garvey and University Negro Improvement Association Papers*. Berkeley: University of California Press, 1987.

Hill, Robert A., and Marcus Garvey. *The Marcus Garvey and Universal Negro Improvement Association Papers*. Berkeley: University of California Press, 1983.

Martin, Tony. *Race First: The Ideological and Organizational Struggles of Marcus Garvey and the Universal Negro Improvement Association*. Westport, CT: Greenwood, 1976.

O'Reilly, Kenneth, and David Gallen. *Black Americans: The FBI Files*. New York: Carroll and Graf, 1994.

Van Deburg, William L. *Modern Black Nationalism: From Marcus Garvey to Louis Farrakhan*. New York: New York University Press, 1997.

Walker, Clement Earl. *Deromanticizing Black History: Critical Essays and Reappraisals*. Knoxville: University of Tennessee Press, 1991.

11

Franklin D. Roosevelt had knowledge of an impending Japanese attack and used Pearl Harbor as an excuse to spur American entry into World War II.

PRO   Rolando Avila
CON   Paul W. Doerr

PRO

As President Franklin Delano Roosevelt observed the unfolding events in Europe in the late 1930s, he became convinced that Adolf Hitler was dangerous and that U.S. assistance was crucial to defend Europe against a Nazi conquest. However, Roosevelt was constrained by the overwhelming isolationism that took hold in the United States after World War I. Working within this political reality, Roosevelt gradually increased U.S. involvement. He supplied the Allies with war materiel, and he began making preparations for U.S. military involvement. But, even after World War II broke out in Europe, the majority of Americans remained strongly opposed to direct U.S. involvement in the fighting. This fact ultimately convinced Roosevelt that the only way he could ever win over public opinion would be if the United States were attacked. Consequently, Roosevelt deliberately provoked Germany, Italy, and Japan, and when the attack on Pearl Harbor occurred, it did not come as a surprise to Roosevelt. In fact, Roosevelt was aware of an impending Japanese attack and welcomed it with genuine mournfulness, because, although he regretted the U.S. sacrifice of lives, he believed it served the greater good in rallying support for U.S. entry into World War II.

Hitler was legitimately elected to German high office in 1933. However, only one year later, he became a dictator and began a small campaign of territorial expansion by reclaimed Alsace-Lorraine from France. Then, in 1939, Hitler initiated a much more aggressive campaign. He invaded Poland, which began World War II when Britain and France declared war on Germany. By 1940, Germany's Blitzkrieg (lightning war) had led to the German occupation of Scandinavia, Belgium, Holland, and a part of France. Furthermore, Hitler had forced British citizens into the subways as he bombed Britain by air. British Prime Minister Winston Churchill asked Roosevelt for help, but even though Roosevelt agreed with Churchill that Hitler posed an obvious threat to the security of the entire world and could not be allowed to go on unchecked, Roosevelt

241

explained that his hands were tied. Unlike many Americans, Roosevelt could see the bigger picture. First, the United States could not afford to lose its allies. Second, if U.S. allies fell, what would stop Hitler from setting his sights on the Americas? He regretted not being able to send troops to help halt Hitler's advance, but, much to his chagrin, the popular political sentiment was entrenched in isolationist ideology, which forced him to comply with the official policy of U.S. neutrality.

The Great Depression, which occurred after World War I, made it easy for Americans to ignore events in the rest of the world. Many Americans were struggling economically and could only see their own problems that they faced on a day-to-day basis. Isolationists believed that U.S. involvement in World War I had been a mistake, and they preached against repeating the same mistake of getting involved in foreign wars that cost much more than they gained for Americans. Instead, many argued, it was far preferable to allow the Atlantic and the Pacific Oceans to isolate the United States from European and Asian wars. Perhaps the most notable sign of the great influence of these isolationist tendencies came at the end of World War I. President Woodrow Wilson's Treaty of Versailles was intended to bring an official end to the Great War, but the U.S. Senate rejected the treaty because many Americans felt that provisions in it made it possible for the United States to be drawn into another European war. Wilson took his case to the American people, delivering 40 speeches in 22 days, but his efforts, which brought on exhaustion and a stroke, were to no avail. Isolationism had taken hold and remained an indomitable force in American politics until the Japanese attack on Pearl Harbor.

Roosevelt was well aware of the seemingly impenetrable wall he faced regarding isolationism. It was under these circumstances that he had to work his political adroitness to do what he felt needed to be done. In fact, several historians have accused Roosevelt of utilizing deceitful tactics to increase gradually U.S. involvement in the war. For example, the U.S. Congress had passed a series of neutrality laws in the 1930s to keep the United States out of war, but Roosevelt convinced Congress to revise the 1935 Neutrality Act, which prohibited loans and the sale of weapons to warring nations. Roosevelt made a strong case. First, he argued that the revision was the best way to keep the United States out of the war, because Britain and France would be able to win the war without U.S. troops. Second, the sale of weapons would be beneficial to U.S. businesses struggling with the depression, because the weapons would only be sold on a "cash and carry" basis. In other words, all weapons had to be paid in cash up front. Third, the responsibility for the transport of the weapons would fall squarely on the shoulders of the buyers, which meant that U.S. ships would not run the danger of transport.

In 1940 Roosevelt convinced Congress to pass the first peacetime military draft in U.S. history and urged them to increase the defense budget. Roosevelt

argued that with World War II raging in Europe it was a good idea for the United States to be able to defend itself. However, historians who accuse Roosevelt of "duplicity" write that these troops were not intended as only a defensive force. Instead, the raising of these troops was part of Roosevelt's ultimate plan to send them to Europe. Also, by executive order, without congressional approval, Roosevelt transferred 50 old U.S. destroyers to Britain in exchange for 99-year leases on eight British naval and air bases in Newfoundland and Bermuda. Again, Roosevelt claimed that the transfer was merely a defensive measure, because the British needed the destroyers to combat German submarine warfare in the North Atlantic, and this was the best way to ensure British victory without sending U.S. troops. Also, the new bases acquired by the United States would give the United States greater security.

Isolationist opposition to Roosevelt's recent policies during his 1940 reelection campaign made it clear to him that he had still not won over enough popular support for his agenda. Roosevelt's critics accused him of plotting to take the United States into an unwanted war. The aversion for war was so strong, in fact, that at one point he came within four percentage points of Wendell Lewis Willkie, his opponent in the presidential election. In order to regain his lead in the polls, Roosevelt, choosing his words carefully, reassured the American people that he would not send American troops to any foreign war. This promise helped him regain the lead and he was reelected. Historians, however, have analyzed the craftiness of his words during his 1940 speeches. Some scholars have concluded that what Roosevelt did not say about war was much more important that what he did say. For instance, Roosevelt said that he would keep the United States out of foreign wars, but he did not explain that if the United States was attacked he would no longer consider the war foreign.

Still drawing the United States closer into war by degrees, Roosevelt explained to the American people in a radio fireside chat that his priority was to keep the United States out of the war. But, in order to do this and ensure a British victory, the United States had to do her part by being the "great arsenal of Democracy." In other words, it would now be necessary to give unrestricted aid to Britain, who could no longer afford to pay for the much-needed supplies. So, the "cash and carry" program was replaced by the "lend-lease" program, which no longer required payment up front. Instead, under this new program, Britain promised to return or replace the borrowed equipment and supplies after the war. Roosevelt explained it in a press conference as an analogy. He said that if a neighbor's house were burning it would be unwise for you to charge the neighbor for the use of your hose. Instead, it would be wiser to allow the neighbor to borrow the hose and save his house and yours from the threat. If neutrality had been debatable before the lend-lease program, it certainly no longer was. The United States was blatantly aiding one side of the war with no more plausible excuse than it merely being a business transaction under the previous "cash and carry" program.

Some scholars have claimed that Churchill lured, or manipulated, Roosevelt into fighting. In 1941 Churchill explained to Roosevelt that American troops would be much more valuable to him in the fight than any equipment he could send. In fact, Churchill kept urging Roosevelt to send troops. If troops were not sent, he insisted, there was a real possibility that all could be lost. In August 1941 Roosevelt and Churchill had a secret meeting on the ship *The Prince of Wales* in Argentia Bay, Newfoundland, where, certainly, Churchill continued his petitions for help from his friend. Although Roosevelt kept the details of the meeting a secret, Churchill later disclosed some of the details to his countrymen when he spoke before the House of Commons. According to Churchill, Roosevelt had pledged to help Britain in the fight at the secret meeting.

However, the isolationist influence was still very strong in America. So, Roosevelt began to try to find a way into the war. For example, he ordered the navy to escort U.S. and allied ships and shoot German and Italian ships on sight. Roosevelt's undeclared war waged on the high seas, however, failed to push Congress to the formal declaration of war he sought. So, Roosevelt came to the sad realization that the only way that poplar opinion would support entering the war was if the United States were attacked. It was necessary for an enemy to strike first, and Roosevelt continued taking steps to increase tensions between the United States and Germany. Roosevelt froze German assets, occupied Iceland with American troops, and began aggressively attacking German submarines. But Germany did not declare war on the United States or launch a massive attack.

So, Roosevelt turned his sights on Japan. At Roosevelt's request, Commander Arthur McCollum drew up an eight-point plan to provoke Japan to "commit the first overt act." Taking ideas from the plan, Roosevelt froze Japanese assets, closed the Panama Canal to Japanese shipping, began making threats to be carried out if Japan did not stop pursuing a Pacific empire, and imposed an embargo. Without excusing the atrocity committed by Japan at Pearl Harbor, arguably, more than any other measure, the embargo prompted Japan to attack the United States. The embargo left Japan with the choice of attacking while it relied on dwindling war materiel or waiting to attack after it was out of the resources and could, therefore, be easily crushed.

After all, Roosevelt's new trade policy categorized Japan as an enemy of the United States. While the United States openly sent free supplies to its European allies under the lend-lease program, Roosevelt refused to sell scrap metals, petroleum, and other products and equipment to Japan that it desperately needed. In fact, the embargo damaged U.S.-Japanese relations to such an extent that many American government officials believed the United States was deliberately provoking an imminent attack. Behind closed doors, Roosevelt mentioned to his cabinet that he was seeking a way to maneuver Japan to fire the first shot. Secretary of War Henry L. Stimson recorded that Roosevelt's dilemma was how to accomplish this without allowing too much damage to the United States. The

U.S. navy was tracking Japanese ships all over the world, and six days before the attack on Pearl Harbor, Roosevelt ordered Admiral Thomas C. Hart, the naval commander in the Philippines, to send three small boats into the path of a Japanese fleet in an effort to provoke a fight. Some historians have speculated that Roosevelt sought to sacrifice the small fleet in order to save Pearl Harbor from attack, but his plan failed, because the small boats were unable to reach the advancing Japanese fleet.

As the drums of war began to sound, Admiral J. O. Richardson, commander of the Pacific Fleet, flew to Washington to explain to Roosevelt that the concentration of ships in Hawaii was not a good military strategy. Richardson suggested alternative and safer ways of stationing the fleet. However, to Richardson's surprise, Roosevelt disagreed with his expert opinion, and the disagreement turned into a heated argument. Within a short time after the two men met, Richardson was dismissed and replaced by Admiral Husband E. Kimmel. After evaluating the harbor, Kimmel also informed Roosevelt of its weaknesses, but Roosevelt was adamant about his decision to place the ships where he wanted them. In fact, more than 90 U.S. ships were stationed at Pearl Harbor. But, mysteriously, none of the U.S. aircraft carriers were in the harbor on December 7. Why? Was it a coincidence? They had all been sent on far away missions, and, consequently, all remained in tack to fight another day. The USS *Saratoga* was off the coast of San Diego, California. The USS *Enterprise* was at Wake Island, and the USS *Lexington* was sailing toward Midway Island. Japan may have very easily seen the concentration of U.S. ships as a window of opportunity that was too good to pass up. More cynical writers have asserted that Roosevelt used the concentration of ships as bait to entrap Japan.

On the Sunday morning of December 7, 1941, Pearl Harbor was not on alert nor was it prepared with the necessary equipment to defend itself. For example, the commanders had requested more airplanes to better defend themselves and scout the area around the islands, but their requests were repeatedly denied. So, Japan did, in fact, launch a surprise attack on the largest U.S. naval base in the Pacific. It was a surprise for the Hawaiian residents and U.S. military presence in the harbor. Before dawn a wave of 183 Japanese torpedo planes attacked the harbor. Half an hour later, a second wave of 167 planes assailed U.S. ships that lay side by side anchored in the harbor. Explosions, fire, and smoke filled the air. Ships were leaking oil that burst into flames and badly burnt sailors who jumped in the water to escape sinking ships. The Oahu airfield was also hit. Many pilots were shot as they tried to board the planes. Death and destruction rained down on Hawaii for about two hours. Twenty-one ships were sunk or damaged, about 3,000 aircraft were destroyed, about 2,000 people were wounded, and over 2,400 people were killed. On the USS *Arizona* alone, 1,177 sailors died as they went down with the ship.

Roosevelt's defenders have denied a conspiracy. These historians have detailed the horrible tragedy at Pearl Harbor and have argued that Roosevelt

Battleships (right to left) USS *Arizona*, *Tennessee*, and *West Virginia* burn after being bombed and torpedoed by the Japanese at Pearl Harbor on December 7, 1941. (Library of Congress)

would not have been capable of such inhumanity. Also, they have claimed that Roosevelt loved the navy, and he would never have allowed it to be destroyed. However, these historians have failed to take into account that no man has ever known the future with absolute accuracy. Certainly, Roosevelt had no way of knowing the extent of the damage that would occur with an attack on the fleet. Perhaps, he had hoped for and expected far less damage. In any case, the next day, Roosevelt went before Congress and asked for a declaration of war in retaliation for the "surprise attack." The American public was stunned and outraged. Under the circumstances, Congress had no choice but to grant his request. From hence forth, "Remember Pearl Harbor" became a rallying call for all Americans.

Yes, Pearl Harbor, Hawaii, was caught by surprise, but Washington, D.C., was not. There were many warnings of the attack—warnings that were not shared with Pearl Harbor high command—which could have prevented or at least minimized the tragedy. By August 1940, Lieutenant Colonel William F. Friedman had secretly cracked "Purple," Japan's highest diplomatic correspondence code using a decryption system nicknamed "Magic." Immediately, the U.S. military began monitoring Japanese coded communications. Months before the attack, for example, military intelligence compiled a series of messages, which requested a mapping (or plotting) of the harbor into sectors, the location of ships, and ship movements from Japanese operatives. On November 14, 1941, the intercepted message informed operatives that U.S.-Japanese relations were at an all-time low

and requested two reports weekly detailing information about all ship movements in the harbor. Nagao Kita, the Japanese consul general, responded to the requests from Tokyo on a regular basis. Shortly before the attack, Japanese agents were ordered to destroy all their records and decoding machines, but, astonishingly, Pearl Harbor high command was left in the dark about all this.

Why? Well, some writers have blamed the military chiefs, who made maintaining Magic a secret a top priority. Military chiefs argued that maintaining the secret that the code had been broken was worth even more than Hawaii! Interestingly, many command posts all over the world were supplied with machines for decoding Purple, but Pearl Harbor was never given one. In fact, Pearl Harbor high command asked for a machine to decode Purple, but the request was denied. So, Pearl Harbor was left at the mercy of Washington for alerts. Some historians have exonerated the military chiefs of a calculated cold-blooded maneuver that led to the sacrifice of Americans and pointed rather to the bureaucratic mess that led to ineffective communication of vital information. Roberta Wohlstetter's *Pearl Harbor, Warning and Decision* (1962) supports this view of communication confusion. The entire process of intercepting, compiling, translating, and communicating important information surely did involved bureaucracy. But, in spite of the bureaucratic confusion that existed, enough critical messages got through the mess to Washington and Roosevelt to sound the alarm.

In October 1941 intelligence officers Captain Alan Kirk and Captain H. D. Bode found out by reading decoded messages that the Japanese were plotting to bomb Pearl Harbor. They went through the required levels of command in order to relay the message to Hawaii. Unfortunately, Rear Admiral Richmond Kelly Turner, their senior officer, declined their request, terminated their intelligence work, and reassigned both men to other duties. In December 1941 three Navy junior officers received attack information and they attempted to inform Pearl Harbor, but their messages were censored. Unfortunately, the censored records led the National Security Agency to conclude in 2008 that these "winds" messages did not reach Washington before the Pearl Harbor attack, and, therefore, Washington was not guilty of sitting on this particular warning. Instead, the real story is that the "winds" messages did not reach Washington or Pearl Harbor before the attack because they were not allowed to do so. Much importance has been placed on Purple. But there were many other sources of information that pointed to a Japanese attack on Pearl Harbor. For example, an intercepted message alerted the military that Tokyo would be using the radio to transmit an "East wind, rain" message. The message would be transmitted three times at the end of a weather report. "East wind" stood for the United States and "rain" signified a diplomatic split. In other words, the transmission of this embedded message would signal to all Japan that the war against the United States had begun. Just as predicted, the message was delivered shortly before the attack on Pearl Harbor.

There were also other warning signs. In early January 1941, a navy memorandum stated that an attack on Pearl Harbor was extremely possible if war broke out. On January 27, 1941, American ambassador to Japan Joseph Grew sent an urgent message to Washington that he had reliable information that in the event of war Japan intended to attack Pearl Harbor, but Pearl Harbor had not been informed. On November 27, the Navy and War Departments issued a warning that Japanese aggression was imminent. On December 6, a Japanese coded message revealed that Japan had reached the conclusion that the United States was asking for a fight. However, none of this information was ever passed on to the Pearl Harbor high command, and these two specific messages would become a special focus for later investigations.

Unfortunately, after years of investigation, the U.S. government pinned the guilt for the disaster on Pearl Harbor's commanders Admiral Husband E. Kimmel and Walter C. Short, who served as scapegoats for more than 50 years. After all, it was reasoned, they were directly charged with the protection of Pearl Harbor. However, on October 30, 2000, Congress expunged the guilty records of both commanders and posthumously gave them back their highest ranks, because, in the opinion of Congress, they were not given critical information. Some critics of the congressional action are upset because they feel that Congress had overstepped its bounds by trying to change history with legislation. Other critics have contended that the exoneration of Kimmel and Short has had the undesirable effect of shifting the blame on Roosevelt and his subordinate Chief of Staff General George C. Marshall. But, conspiracy theories surrounding Roosevelt's guilt were not new in 2000. They have been around ever since the attack in 1941. And they remain a matter of debate throughout the successive high-level military inquires (the Roberts Commission, the Hart Inquiry, the Army Pearl Harbor Board, the Navy Court of Inquiry, the Clarke Inquiry, the Clausen Investigation, and the Hewitt Inquiry). In 1946 Congress also conducted its own investigation in an effort to answer the all-consuming question of who was to blame for the tragedy.

In 1945 journalist John Chamberlain introduced the conspiracy theory to a wide audience when he wrote an influential article in *Life* magazine in which he accused Roosevelt of knowing about the attack and allowing it to happen so that the United States would be drawn into World War II. The popular obsession with the topic has shown no signs of abating. In the December 15, 1999, issue of the *New York Times*, Richard Bernstein again raised the question of foreknowledge in an article titled "On Dec. 7, Did We Know We Knew?" And, again on April 18, 2004, another *New York Times* writer Joseph E. Persico wrote an article titled "What Did He Know, and When?" By the date of this writing, conspiracy books, Web sites, and pod casts were dedicated to the "Back Door to War" theory.

Some studies have defended Roosevelt at every turn. The chief weakness of most of these works, however, has been that in attempting to address conspiracy charges, they too often have written off Roosevelt's masterly crafted duplicity

as mere blunders of a flawed hero. In other words, they have claimed that the best way to describe Roosevelt's role in the attack on Pearl Harbor was not conspiracy, but, instead, well-intentioned stupidity. We are led to believe that Roosevelt was an incompetent fool before the attack, but he blossomed into a political genius only after the attack. The book by Gordon W. Prange's *At Dawn We Slept: The Untold Story of Pearl Harbor* (1981) is an important work that holds to the flawed hero explanation of events. On the other extreme of the spectrum, some works have appeared that have gone too far by making some unsupportable and even outlandish accusations about Roosevelt's role that have appeared more like tabloid front pages than historical writing. For example, Mark Emerson Willey's *Pearl Harbor: Mother of All Conspiracies* (2000) claims that Roosevelt, a communist according to Willey, always worked for Stalin's best interest and brought the United States into war in order to lead the Russians against Hitler. But, most astonishingly, Roosevelt's major goal, according to Willey, was to build a communist world empire.

In 1948 Charles Austin Beard, a highly respected historian, established Roosevelt's "Back Door to War" theory as a legitimate matter for scholarly debate with the publication of his first book on the topic titled *President Roosevelt and the Coming of the War, 1941*. Since then, many writers have continued to expand on the ever-present Roosevelt's "Back Door to War" theme. One valuable study is *Back Door to War: The Roosevelt Foreign Policy, 1933–1941* (1952) by Charles C. Tansill, which has become the classic account. Also of great value is *The Final Secret: The Washington Contribution to the Japanese Attack* (1954) by Rear Admiral Robert A. Theobald, who served in the military during the Pearl Harbor attack and was present at one of the investigative hearings. One of the most thorough analyses of the events is *Infamy: Pearl Harbor and Its Aftermath* (1982) by John Toland, which makes good use of newly declassified documents and testimonies available to him by the early 1980s. *Betrayal at Pearl Harbor: How Churchill Lured Roosevelt into World War II* (1991) by James Rusbrider and Eric Nave documents the great influence that Churchill held over Roosevelt and even suggests that Roosevelt was tricked into doing Churchill's bidding. And, finally, *Day of Deceit: The Truth about FDR and Pearl Harbor* (1999) by Robert B. Stinnett, revitalizes the conspiracy theory by making use of newly declassified documents in the 1990s.

Astonishingly, even with the weight of the evidence of Roosevelt's knowledge of the attack, some skeptics have still charged that the smoking gun has remained elusive. They have held to the position that all of the before-mentioned facts are just circumstantial. But, common sense would dictate that Washington, D.C., would not have kept the information a secret from the chief executive. Roosevelt's leadership style was a hands-on approach. There is no way that Roosevelt did not know of the attack. Magic was delivered to Roosevelt's desk on a regular basis. He read many of the intercepted messages for

himself. On one occasion, after reading an intercept, he was heard concluding that Japan would soon be at war with the United States. Also, a lot of other information, which came before the president from undercover sources, alerted him to the danger. Spies abounded during this time of crises. In early 1941 Congressman Martin Dies came into the possession of a Japanese strategic map that was prepared by the Japanese Imperial Military Intelligence Department, which showed that Japan clearly intended to launch an attack on Pearl Harbor. Dies contacted Secretary of State Cordell with the warning. Cordell told Roosevelt, and Roosevelt ordered Cordell to tell Dies to keep the matter a secret. Cordell informed Dies that Roosevelt did not wish for the information to leak to the news media, because he considered it an essential secret of national defense. In the summer of 1941, Dusko Popov, a Yugoslav spy working for the United States, informed the Federal Bureau of Investigation (FBI) in New York that German agents were helping Japan to map out Pearl Harbor into grids for a surprise Japanese air strike. J. Edgar Hoover informed Roosevelt of the information, but Roosevelt ordered him to keep it a secret. After the war, Hoover angrily recalled how Roosevelt had told him to just leave the information in his capable hands. In November 1941 Kilsoo Haan, of the Sino-Korean People's League, discovered from the Korean underground that Japan would make an attack on Pearl Harbor before Christmas. Haan told Iowa Senator Guy Gillette, who in turn informed Roosevelt.

Some writers have argued that Roosevelt would not have provoked a war with Japan if what he really wanted was to help Britain fight against Germany. These writers fail to take into account that at the time of the Pearl Harbor attack, Japan had formed an alliance with Germany and Italy, and declaring war

---

Robert B. Stinnett and the Pearl Harbor Conspiracy

One of the main supporters of the view that British and U.S. authorities knew of the impending attack on Pearl Harbor and failed to prevent it as it would bring the United States into World War II is Robert B. Stinnett, a former U.S. sailor who served as a naval photographer in World War II, serving in the same unit as George H. W. Bush, which led to his book, *George Bush: His World War II Years* (1992).

Stinnett was born on March 31, 1924, at Oakland, California, the oldest child of Curtis and Margaret Stinnett, and as a teenager, he started taking photographs for the *Oakland Tribune*. In World War II, he served on the aircraft carrier *San Jacinto* and then returned to Oakland, becoming a photographer and journalist for the newspaper.

It was while reading Gordon W. Prange's book *At Dawn We Slept* that Stinnett first came across allegations on U.S. foreknowledge, and he started writing about it for his newspaper, and then for his book *Day of Deceit: The Truth about FDR and Pearl Harbor* (1999).

against Japan would inevitably lead to a war declaration from Germany. And, not surprisingly, Germany declared war against the United States just one day after the United States declared war against Japan. During the war, tens of thousands of Japanese Americans were rounded up as security risks and sent to internment camps during World War II. It was a sad chapter in U.S. history. But, even more shocking was the fact that Roosevelt ordered the Census Bureau to compile a list of Japanese American names and addresses before the attack on Pearl Harbor even occurred. In 2007 researchers in the National Archives uncovered an important memorandum from William J. Donovan, director of the Office of Strategic Services, to Roosevelt. The memorandum, dated November 13, 1941, supports the charge that Roosevelt backed Japan into a corner by ultimatums and deadlines. Donovan wrote Roosevelt that it was plain to see that, if Japan waited too long to attack the United States, it would be "strangled" by a lack of resources. However, Donovan reassured Roosevelt that Japan would, therefore, have no choice but to attack. Without a doubt, Roosevelt knew of the coming attack on Pearl Harbor and he capitalized on it to bring the United States into World War II.

References and Further Reading

Bachrach, Deborah. *Pearl Harbor: Opposing Viewpoints*. San Diego, CA: Greenhaven, 1989.

Baker, Kevin. "The Guilt Dogging the Greatest Generation." *New York Times* (November 12, 2000): WK16.

Beard, Charles Austin. *President Roosevelt and the Coming of the War, 1941*. New Haven, CT: Yale University Press, 1948.

Buchanan, Patrick J. *A Republic, Not an Empire: Reclaiming America's Destiny*. Washington, DC: Regnery, 1999.

Morgenstern, George. *Pearl Harbor: The Story of the Secret War*. New York: Devin Adair, 1947.

Perloff, James. "Pearl Harbor: Hawaii was Surprised, Washington Was Not." *New American* (December 8, 1986): 25–36.

Prange, Gordon W., with Donald M. Goldstein and Katherine V. Dillon. *At Dawn We Slept: The Untold Story of Pearl Harbor*. New York: Penguin, 1982.

Roberts, Sam. "Report Debunks Theory that the United States Heard a Coded Warning about Pearl Harbor." *New York Times* (December 7, 2008): 30.

Rusbridger, James, and Eric Nave. *Betrayal at Pearl Harbor: How Churchill Lured Roosevelt into World War II*. New York: Summit Books, 1991.

Stinnett, Robert B. *Day of Deceit: The Truth about FDR and Pearl Harbor*. New York: Free Press, 1999.

Tansill, Charles Callan. *Back Door to War: The Roosevelt Foreign Policy, 1933–1941*. Westport, CT: Greenwood, 1952.

Theobald, Robert A. *The Final Secret of Pearl Harbor: The Washington Contribution to the Japanese Attack*. New York: Devin-Adair, 1954.

Toland, John. *Infamy: Pearl Harbor and its Aftermath*. Garden City, NY: Doubleday, 1982.

Willey, Mark Emerson. *Pearl Harbor: Mother of All Conspiracies*. Philadelphia: Xlibris, 2000.

Wohlstetter, Roberta. *Pearl Harbor: Warning and Decision*. Stanford, CA: Stanford University Press, 1962.

CON

On the morning of Sunday December 7, 1941, beginning at 7:53 A.M., carrier-borne aircraft of the Japanese Imperial fleet unleashed a devastating surprise attack against the U.S. Pacific Fleet at Pearl Harbor on the Hawaiian island of Oahu. Six American battleships were sunk and two others damaged along with 10 additional vessels either sunk or damaged. The Japanese destroyed 164 American aircraft and damaged another 128 on the Oahu island airfields. The American death toll amounted to 2,403 military personnel and civilians, plus 1,178 wounded. The Japanese lost 29 aircraft, 5 midget submarines, and 1 fleet submarine. The United States declared war on Japan the next day. Hitler's Germany declared war on the United States on December 11.

Recriminations and finger pointing in the United States began almost immediately. No fewer than seven official investigations and commissions of inquiry were carried out during the period from 1942 to 1946. The Pearl Harbor conspiracy theory surfaced soon after the attack and won some converts in the 1940s and 1950s. Despite criticism from historians, it has enjoyed a stubborn longevity ever since, testifying to the enduring popularity of conspiracy theories. Books articulating the conspiracy theory emerge every now and then, and two of the more popular recent volumes will be examined below. The emergence of the Internet in the 1990s gave conspiracy theorists another lease on life and some more converts among the uninformed. The theory has a number of variations that will be examined here, but the basics are clear. Allegedly, President Roosevelt deliberately exposed the U.S. Pacific Fleet to attack and destruction by the Japanese. Why would a president do such a thing? Roosevelt's sympathy for the British in their war against the Axis powers (Germany, Italy, and Japan) is supposedly the key. Roosevelt had allegedly long been itching to declare war on the Axis powers but had been stymied by isolationist sentiment in the United States. According to the conspiracy theorists, also known as revisionists, Roosevelt hoped that public outrage over the attack would allow him to overcome

public opinion and take the United States into the war via "the back door." Supporters of the Pearl Harbor conspiracy theory have usually cited allegations that Roosevelt had advanced warning of the attack, but withheld this crucial information from Admiral Husband E. Kimmel, commander in chief of the U.S. Pacific Fleet and Lieutenant General Walter C. Short, commanding general of the Hawaiian defenses, thus ensuring that the Japanese would attain surprise. The proof for this outrageous theory is entirely unconvincing and has never won the support of any reputable historian. Not a single government document has ever come to light that supports the claims of conspiracy theorists.

The disaster at Pearl Harbor was not the result of a sinister Machiavellian conspiracy hatched by President Roosevelt. Instead the American defeat resulted from meticulous Japanese planning and American misjudgments about Japanese intentions and capabilities. Although the Americans did anticipate a Japanese attack, they believed it would come elsewhere in the Pacific. The main threat to the Pearl Harbor base was believed to be from saboteurs, not a long-range attack by carrier-borne aircraft, which the Americans considered was beyond Japanese capabilities. General Short therefore ordered that aircraft be grouped together on airfields where they could be more easily guarded, which unfortunately made them perfect targets for attacking Japanese aircraft. Antitorpedo nets had not been provided for the battleships in the harbor because it was feared they would impede the mobility of the ships in an emergency. American radar operators on Oahu were poorly trained and inexperienced and mistook incoming Japanese aircraft on the morning of December 7 for an expected flight of American B-17 bombers. The Americans also believed that the waters of Pearl Harbor were too shallow for effective use of torpedoes. The British had great success in adapting their torpedoes for an attack on the Italian fleet at Taranto harbor on November 11, 1940, but Pearl Harbor was substantially shallower than Taranto. In fact the Japanese did not succeed in altering their torpedoes for use in Pearl Harbor until early November 1941, and then kept the new design strictly secret. On the few occasions when an attack on the Hawaiian Islands was discussed, American planners naturally anticipated an attack from the Japanese-held Marshall Islands, to the southeast of Hawaii. Instead the Japanese attack came from the north.

Charges that Roosevelt deliberately exposed the Pacific Fleet to Japanese attack also make little sense on a strategic level. The allegation that Roosevelt threw away the pride of the Pacific Fleet on a wild gamble to gain a declaration of war from Congress is simply too far-fetched. Why would a president sacrifice the core of his country's navy immediately prior to a war? Conspiracy theorists often point to the absence of U.S. fleet carriers and cruisers from Pearl Harbor as "proof" that Roosevelt knew what was coming and decided instead to sacrifice "obsolete" battleships. However, the importance of aircraft carriers in the Pacific was not clear until early 1942. Roosevelt was an orthodox naval thinker who thought that battleships would play a key role in the next war.

Roosevelt, a former assistant secretary of the navy, was an ardent fan of the U.S. navy, as is apparent from every biography written about his life. A visitor to his home, located north of New York City and preserved as a national memorial, will find it full of naval memorabilia, showing his fondness and regard for the navy. One could speculate that militarily it would have made more sense to move the U.S. battleships out of Pearl Harbor and then use them to unleash a surprise ambush on Japanese carriers once their aircraft had taken off to attack an empty harbor. It is also worth remembering that at some point historians must produce solid evidence for their theories and not just continually indulge in speculation and fantasies.

Roosevelt simply did not want war with anyone in 1941; he was convinced that U.S. defenses were too weak and that more time was needed to prepare the country for war. Roosevelt regarded Germany as the major threat and believed that a war in the Pacific would completely distract the United States from focusing on Germany. As well, the Pearl Harbor conspiracy theory requires the involvement not only of Roosevelt, but all his senior military and political advisers and cabinet ministers, practically all of whom incidentally wanted to avoid war with Japan, along with hundreds of lower ranking personnel. All these individuals would have to operate together to make the conspiracy work, without a single one of them spilling the beans either in the lead up to the attack or in the years thereafter. Conspiracy theorists usually see Kimmel and Short as the patsies for the Pearl Harbor disaster. Logically it would have made more sense to have them in on the conspiracy since that way Roosevelt could be sure that they would not trigger an alert that might stop the attack, but this is not an aspect of the conspiracy theory that revisionists want to deal with. Revisionists also frequently argue that the Japanese would have turned back if they had been detected by the Americans prior to the attack, but that is simply not true. The Japanese fully expected to fight their way into Pearl Harbor and would not have given up their mission if they had been detected.

Furthermore, conspiracy theorists are not clear on exactly what Roosevelt hoped to accomplish. Certainly a war with Japan would result, but what about Germany? It was by no means clear that Roosevelt could get a declaration of war against Germany through Congress in the days after the attack. Hitler relieved him of that task by declaring war on the United States, an unexpected act. The Tripartite Alliance, Germany's alliance with Japan and Italy, was a defensive alliance. Since Japan had attacked the United States, Hitler technically had no obligation to declare war on the United States and could easily have weaseled out. However, Hitler decided to stand by his ally even though he had never before kept his word on anything. How was Roosevelt supposed to know that Hitler would declare war on the United States in the aftermath of Pearl Harbor?

It is also clear that U.S. intelligence was snowed under by a veritable blizzard of information as the Japanese moved toward war in the fall of 1941.

American intelligence agencies were understaffed, overworked, and proved unable to prioritize properly the threats that the United States faced. Different intelligence agencies of the military seldom worked together to coordinate their findings. Another major bottleneck to gaining accurate and timely intelligence was the problem of finding competent translators. As well the Japanese were completely successful in cloaking the Pearl Harbor operation in secrecy. U.S. intelligence in early December indicated that the Japanese were moving ships and troop transports along the Chinese coast in the direction of Indochina for anticipated attacks against the Philippines, Thailand, or the Dutch East Indies. Intelligence also indicated that Japanese aircraft carriers were in home waters and not at sea (Wohlstetter 1962: 382–96).

President Franklin D. Roosevelt signs the declaration of war against Japan on December 8, 1941. (National Archives)

What was the state of U.S. intelligence gathering in late 1941? Since 1935 a U.S. naval intelligence unit operating under the code name Magic had been decrypting Japanese military codes with some degree of success. A major breakthrough was scored in 1939 when the Japanese diplomatic code, Purple, was broken. This enabled Americans to read telegrams from Tokyo to Japanese embassies around the world. Readers should remember however that Purple was only the diplomatic and not the military code. Progress had also been made in decoding lower-level Japanese codes such as the J-19 code used by the Japanese consul in Honolulu. The Americans were also successful in locating Japanese ships through an analysis of radio traffic, although the contents of coded radio messages themselves could not be read. The Japanese could also throw the Americans off temporarily by changing their radio codes, which they did on November 1 and December 1. The keys for the J-19 code changed frequently, and decryptionists needed a fair volume of messages to break them. Translations and transmissions to Washington required more time. Then the intelligence had to be correctly interpreted, an activity fraught with the potential for mistakes. Breaking codes does not necessarily mean that an intelligence service has an open door into the mind of a potential enemy.

Generally speaking, U.S. intelligence, by the fall of 1941, had a picture of Japan moving steadily toward hostile action, but no specific warning of an

attack on Pearl Harbor. On November 27 the office of the chief of naval operations issued a warning that an aggressive move from the Japanese could be expected in the next few days. The warning was issued to U.S. military commanders across the Pacific, including Kimmel and Short. But an attack on Pearl Harbor, outside of local sabotage, was not foreseen. Attacks on the Philippines and the Dutch East Indies were predicted and did materialize. Another intelligence thread suggested a Japanese attack on the USSR at a time when the Soviets were desperately fighting the Germans in front of Moscow. Why was an attack on Pearl Harbor so hard to foresee? U.S. military officials had pretty much made up their minds as to where the Japanese were going to attack and nothing suggested that the Japanese were about to do anything different.

Over the years Pearl Harbor conspiracy theorists have cited a variety of specific incidents that they think prove the existence of the conspiracy. There is such an endless succession of these events that refuting all of them in a single article is not possible. I have chosen instead to focus on some of the most popular stories cited by conspiracy theorists. Generally speaking, conspiracy theorists adopt a similar approach to "the evidence," seizing upon an isolated incident, blowing it completely out of proportion, taking it out of context, and repeating charges until they assume mythical status in the pantheon of revisionist history. Not one of the stories examined below stands up to historical scrutiny.

The first incident to be examined concerns the role of Dusko Popov, a wealthy, high-living Yugoslavian business lawyer with a German degree. Popov had been recruited in Belgrade by the German secret service, the Abwehr, in 1940. What the Abwehr did not know was that Popov had already been recruited by the British Secret Intelligence Service. Popov therefore went to work for the Abwehr as a British double agent. In June 1941 the Abwehr sent Popov to New York with orders to set up a German spy ring in the United States. Popov was given an extensive questionnaire by the Abwehr to guide his espionage activities. Basically Popov was to try to obtain information on U.S. defense installations and the extent to which the Americans were supplying the British war effort. Using the latest technology at the time, the questionnaire given to Popov was in the form of microdots on four paper documents. Popov informed his British handlers of what he had been instructed to do and supplied them with the questionnaire. Upon arrival in the United States, Popov met briefly with J. Edgar Hoover, head of the Federal Bureau of Investigation (FBI). Hoover refused to take Popov seriously and sent him on his way.

The microdot questionnaire shows that the Abwehr was interested in U.S. installations at Pearl Harbor. The questionnaire also shows that the Abwehr was seriously interested in many other aspects of the U.S. defense establishment. Popov was supposed to gain information on U.S. tank and aircraft production, financial details of American-British cooperation, and tables of organization for U.S. infantry divisions. Popov was even instructed to gain information on the

---

The FBI and Japanese Wiretaps

The FBI had been monitoring the telephone calls made from Hawaii to Japan, and vice versa, and several days before the attack on Pearl Harbor, they had listened to a staffer from the Tokyo newspaper *Yomiuri Shinbun* calling Motokazu Mori (1890–1958), a dentist in Hawaii. Mori's wife was the Honolulu correspondent for the newspaper. The journalist in Tokyo asked Mori about current life in Hawaii and included questions on airplanes, searchlights, the weather, sailors, and also about flowers. Among the reply by Mori was a comment, 'presently, the flowers in bloom are fewest out of the whole year. However the hibiscus and the poinsettia are in bloom now.' The FBI immediately suspected that the description of the flowers was a coded message for something sinister and arrested him. A later investigation found that the description of the flowers did appear in the *Yomiuri Shinbun* on the following morning and Mori said that the description was for some background atmosphere.

Canadian air force. The Abwehr wanted particulars on all ports of assembly for overseas transport in the United States and Canada.

The revisionists, particularly John Toland, have had a field day with the Popov questionnaire. Toland, in his book *Infamy: Pearl Harbor and Its Aftermath* (1982) alleges that the questionnaire was a virtual tipoff, a detailed plan of attack, betraying Japanese intentions to attack Pearl Harbor. Toland also alleges that Hoover, Roosevelt, and the British acted to suppress this information so that the Japanese attack could go ahead. In order to prove his point Toland indulged in some highly selective and deceptive editing. He only quotes from the section of the questionnaire dealing with Pearl Harbor, ignoring all the other information that Popov was to gather and creating a highly misleading and exaggerated impression. There are no detailed plans for an attack at all, and Toland does not even inquire into the extent of Japanese-German intelligence cooperation in 1941.

Furthermore, Popov's information was not taken seriously for the good reason that neither the British nor the Americans could be certain he was a reliable informant. The British intelligence network in Europe had been broken by the Germans in 1940. The British had to be very careful with Popov in case he was a triple agent actually working for the Abwehr (Bratzell et al. 1982; West 1985: 68–81). However, the Popov questionnaire does tell us a great deal about how revisionists approach evidence and the manner in which evidence can be taken out of context and twisted to prove a point. Readers should also be aware of the use of hindsight, a common revisionist tactic. Things that seem obvious to us now may not have been clear at all to people at the time.

Another popular revisionist story, also circulated by Toland, concerns a more immediate warning that the Americans purportedly received just prior to the attack. Basically the allegation is that the Japanese fleet, on its way to Pearl Harbor, in complete defiance of all military sense, broke radio silence. Japanese

transmissions were supposedly picked up by American radio operators. Their warnings were then stifled by superior officers who were presumably part of the conspiracy.

In *Infamy*, Toland interviewed a former U.S. naval radio operator based in San Francisco who claims to have detected radio signals from the approaching Japanese fleet. How San Francisco was able to do this, while radio operators based in Hawaii, much closer to the approaching Japanese fleet, missed the signals is conveniently not explained. Surviving Japanese naval personnel have always testified that their fleet observed the strictest radio silence and that many of their radios had been disabled to prevent accidental usage. Toland's source was discovered living in San Francisco and was interviewed in 1983 by the U.S. navy and other prominent historians. They concluded that this individual's story had no credibility, that he had no access to any firsthand information, that Toland had distorted his account, and that it was impossible for a radio operator in San Francisco to have detected radio signals in the distant Pacific on the frequencies used by the Japanese navy (Coox 1986; Prange 1986: 55–58).

A related story concerns the passenger liner *Lurline*, which allegedly detected Japanese radio signals from ships north and west of Hawaii. When docking at Honolulu the crew supposedly informed navy authorities. The ship then proceeded on to San Francisco where it docked on December 10. A naval officer came aboard and took possession of the ship's log. Since Toland could find neither the Honolulu report nor the *Lurline* log in the navy archives, the implication is that the information was suppressed. This story is not credible simply because the Japanese did not break radio silence. The U.S. navy took possession of the logs of all ships at sea on December 7, so there was nothing sinister about the seizure of the *Lurline*'s log (Prange 1986: 52–55).

Toland also relies on a story told by Captain Johan Ranneft, a Dutch naval attaché stationed in Washington at the time of the attack. Toland claims that Ranneft's diary records instances on December 2 and again on the December 6 in which officials in Washington showed him maps on which the progress of the Japanese fleet toward Pearl Harbor was tracked. Toland claims this as proof that U.S. intelligence knew that the Japanese were about to attack Pearl Harbor and did nothing to prevent it on orders from the top. However, Ranneft's diary refers to two Japanese carriers, while the Pearl Harbor fleet had six carriers, so the Americans could not possibly have identified the Pearl Harbor attack force. As well, Toland appears to have distorted Ranneft's testimony to suggest the carriers were far closer to Pearl Harbor than Ranneft had actually stated in his diary. The two carriers in question were in fact located at Kwajalein Island in the Marshall Islands. The location of these carriers at Kwajalein has long since been known to historians and Toland had discovered nothing new. The Americans assumed the carriers were part of a routine surveillance force. The two carriers played no part in the Pearl Harbor attack, and the Pearl Harbor fleet did not begin

its journey from the Marshall Islands. Other historians have raised serious questions about Ranneft's personal credibility as he admitted leaking Dutch and Swedish patents for an antiaircraft gun to the Americans. He then apparently lied to cover up his indiscretion (Prange 1986: 60–61). The whole story also implies once again that the Japanese broke radio silence, which, as we have seen, did not happen. For a successful attack on Pearl Harbor, the Japanese absolutely had to maintain radio silence. Why then would they jeopardize the success of the operation by talking on open radios across the whole of the Pacific? Toland is an admitted hater of Roosevelt, and his allegations tell us much about the desperation of revisionists to make facts fit a preconceived theory.

Two other stories about early warnings of the Pearl Harbor attack are also frequently circulated by revisionists. The first concerns the story of the "three little ships." On December 1 Roosevelt ordered three small naval vessels to take up positions off the Indochina coast to determine the direction of a possible Japanese invasion fleet. Roosevelt, as mentioned above, was a big fan of the navy and often indulged himself in the minutiae of naval operations. As it turned out, only one vessel, the *Isabel*, had reached its station on December 7. Revisionists charge that the three little ships were actually to serve as bait for a Japanese attack. If they were sunk or attacked, then Roosevelt would have his excuse for war. Again this is the purest fantasy and ignores the fact that an attack on a tiny vessel (one of the three ships displaced only 150 tons) would not provide sufficient grounds for a declaration of war. Roosevelt had ignored far worse attacks on U.S. ships by the Germans in the North Atlantic (Prange 1986: 45–51).

A more serious allegation concerns the so-called bomb plot message. On September 24 the Japanese Naval Intelligence branch requested that the Japanese consulate on Oahu step up surveillance of Pearl Harbor. In particular the consulate was instructed to divide the harbor into grids and then report on which ships were positioned in those grids. As was mentioned above, the Americans could read the consulate codes, but this message was not decoded and translated until October 9, reflecting the backlog that the United States was experiencing in the decoding and translating process. The divided nature of the U.S. intelligence community also ensured that the seriousness of the message was not appreciated. The September 24 message should have triggered suspicions, if only because Oahu was the home of the Pacific Fleet. Senior military officials on Oahu, however, saw nothing new or alarming in the message. They dismissed it as routine Japanese military requests for details on U.S. fleet comings and goings, an attempt to streamline and simplify the intelligence flow, or more evidence of typical Japanese attention to detail. The Japanese were requesting similar information from other consulates regarding harbors such as those at Manila, the Panama Canal, Portland, and San Diego. The September 24 message was just one of a host of reports coming through U.S. intelligence channels. As such it simply failed to provoke any alarm and reflects a common failing in intelligence collection. Coming into possession

of intelligence does not always mean that it will be interpreted properly. Human beings still have to make the correct judgments and are prone to making errors (Prange 1982: 248–57; Prange 1986: 267–84; Wohlstetter 1962: 211–14).

Revisionists have also made much of the so-called winds execute story, another purported tipoff. Upon the outbreak of war, Japanese diplomatic stations abroad were instructed to burn their codebooks, a common security procedure. The Japanese decided they needed a special coded message to signal their diplomatic missions to begin burning their codes. Certain phrases would be included in radio broadcasts from Japan. The phrase "east wind rain" meant that relations with the United States were in danger. "North wind cloudy" meant trouble with the USSR. "West wind rain" meant trouble with Britain. The appropriate phrases would be used in the middle and end of a weather broadcast and repeated twice. Upon receipt of the appropriate "winds" message, the missions would take action. The Japanese sent a message to their diplomatic missions informing them of the procedure. U.S. intelligence was aware of the Japanese plan and then expended much energy looking for the winds execute message, that is the winds message that told diplomatic missions to take action. Contrary to what the revisionists say, the message was never sent and, even if it had been, the information would have been useless. It was obvious to everyone by the end of November that relations between the United States and Japan were in a great deal of trouble (Prange 1982: 358–59).

Revisionists also frequently allege that Churchill cooked up a secret deal with Roosevelt to bring the United States into the war. Suspicion tends to fall on the August 1941 meeting that Roosevelt had with Churchill on board the British battleship *Prince of Wales*, anchored off Argentia, Newfoundland. Some revisionists have claimed that Roosevelt promised to enter the war if the Japanese attacked British possessions in the Pacific. Aside from the fact that not a single document has surfaced supporting such an interpretation, conspiracy theorists totally misrepresent what happened at the meeting. With Britain in a desperate struggle for survival, Churchill was hoping to limit the possibility of a war with Japan. He hoped that a stronger diplomatic stand from the United States might deter Japan, but Churchill sailed home profoundly disappointed. On August 17 the United States issued a note to the Japanese government stating that Washington would defend American lives and interests if threatened by Japanese aggression, which is the plain, obvious duty of any government worth its salt (Prange 1986: 76). The Americans never promised to go to war with the Japanese in defense of British or Dutch imperial interests.

As has been mentioned above, the revisionist view of Pearl Harbor has not found the support of a single reputable historian. John Toland's revisionist account of Pearl Harbor has been comprehensively discredited. Representatives of an older generation of historians who veered off into Pearl Harbor revisionism, such as Harry Elmer Barnes, have also seen their reputations suffer. Barnes in particular also became involved in Holocaust denial (Lipstadt 1993: 67–71).

Most recently Robert Stinnett's *Day of Deceit* (1999), another entry in the revisionist camp, similarly received dismissive reviews in the professional journals, most notably the *Journal of Military History*. Stinnett's knowledge of his subject was simply too flimsy to convince anyone except the diehards. He showed no understanding of how code breaking worked in 1941 or the structure and functioning of the U.S. and Japanese militaries. The possibility that all professional historians are in on Roosevelt's secret is simply too absurd to discuss.

Newcomers to the issue might begin by seeking out some of the more reputable historical studies of the origins of the war in the Pacific, such as Akire Iriye's *The Origins of the Second World War in Asia and the Pacific* (1987). The essays in Richard Overy and Andrew Wheatcroft's *The Road to War* (1999) are illuminating as are the essays in Robert Boyce and Joseph Maiolo's *The Origins of World War Two: The Debate Continues* (2003). These three books will provide curious readers with the basics for understanding the origins of the war in the Pacific, based on the latest reputable historical research, which involves many years of detailed archival research and balanced judgments. Those interested in more detail should consult Jonathan Utley's *Going to War with Japan, 1937–1941* (1985) and Waldo Heinrichs's *Threshold of War* (1990). Roberta Wohlstetter's *Pearl Harbor: Warning and Decision* (1962) is a masterful analysis of the failures and shortcomings of the U.S. intelligence apparatus prior to the attack. The king of Pearl Harbor historiography is undoubtedly Gordon Prange, much of whose work remained unpublished at the time of his death in 1980. Prange's *At Dawn We Slept* (1981) is a detailed account of the attack that also contains a chapter ("Revisionists Revisited") refuting revisionist accounts. A second posthumous book, *Pearl Harbor: The Verdict of History* (1986) contains over 600 pages of detailed rebuttals of the revisionist arguments. Both books were published with the assistance of Donald Goldstein and Katherine Dillon. The fact that so much of Prange's work is focused on the revisionists has led other historians to lament the extent to which the Pearl Harbor research agenda is still being set by obsolete revisionist theories, when other issues still need to be explored (Melosi 1983: 87–103). For example, how did the Pearl Harbor disaster impact Cold War defense planning and anxieties about a future surprise attack? Clearly it is time to ditch Pearl Harbor conspiracy theories and move on.

The events at Pearl Harbor on December 7, 1941, seem so obvious that we are often baffled as to why people at the time did not see the attack coming. We have seen similar reactions to the events of September 11, 2001. Much of the revisionism around Pearl Harbor is fueled by a lingering hatred of Roosevelt and a determination to sully his considerable legacy. Readers should remember that Roosevelt could not predict the future and he could not force the Japanese to do what they did not want to do. The Japanese made their own decisions regarding peace and war. They were not manipulated by a "sinister" Roosevelt sitting in the White House thousands of miles away pulling strings.

References and Further Reading

Boyce, Robert, and Maiolo, Joseph A., eds. *The Origins of World War Two: The Debate Continues*. New York: Palgrave Macmillan, 2003.

Bratzel, John F., and Rout, Leslie B., Jr. "Pearl Harbor, Microdots and J. Edgar Hoover." *American Historical Review* 87 (1982): 1342–1351.

Coox, Alvin D. "Repulsing the Pearl Harbor Revisionists: The State of Present Literature on the Debacle." *Military Affairs* 50, 1 (January 1986): 29–31.

Iriye, Akira. *The Origins of the Second World War in Asia and the Pacific*. New York: Longman, 1987.

Heinrichs, W. *Threshold of War: Franklin D. Roosevelt and American Entry into World War Two*. Oxford: Oxford University Press, 1990 reprint.

Lipstadt, Deborah. *Denying the Holocaust: The Growing Assault on Truth and Memory*. New York: Free Press, 1993.

Melosi, Martin V. "The Triumph of Revisionism: The Pearl Harbor Controversy, 1941–1982." *Public Historian* 5, 2 (Spring 1983): 87–103.

Mintz, Frank P. *Revisionism and the Origins of Pearl Harbor*. Lanham, MD: University Press of America, 1985.

Overy, Richard, and Andrew Wheatcroft. *The Road to War*. London: Penguin, 1999.

Prange, Gordon W., with Donald M. Goldstein and Katherine V. Dillon. *At Dawn We Slept: The Untold Story of Pearl Harbor*. New York: Penguin, 1982.

Prange, Gordon W., with Donald M. Goldstein and Katherine V. Dillon. *Pearl Harbor: The Verdict of History*. New York: McGraw-Hill, 1986.

Rusbridger, James, and Eric Nave. *Betrayal at Pearl Harbor: How Churchill Lured Roosevelt into War*. London: Michael O'Mara Books, 1991.

Stinnett, Robert. *Day of Deceit: The Truth about FDR and Pearl Harbor*. New York: Free Press, 1999.

Toland, John. *Infamy: Pearl Harbor and Its Aftermath*. New York: Doubleday, 1982.

Utley, Jonathan. *Going to War with Japan, 1937–1941*. Knoxville: University of Tennessee Press, 1985.

West, Nigel. *A Thread of Deceit: Espionage Myths of World War Two*. New York: Random House, 1985.

Wohlstetter, Roberta. *Pearl Harbor: Warning and Decision*. Stanford, CA: Stanford University Press, 1962.

12

Alger Hiss's 1950 conviction for espionage was not an example of Cold War hysteria. He was a Soviet spy and deserved his punishment.

PRO   Jeffrey H. Bloodworth
CON   Annessa Babic

PRO

For more than a generation the controversy surrounding Alger Hiss has served more as political Rorschach test than search for historical fact. To his defenders, anticommunist zealots fabricated and exaggerated evidence to frame Hiss and defame the left. To opponents, Hiss represented the American left's naivete and concomitant collusion with global communism. Putting the partisan and ideological bromides of the past aside, new evidence and scholarship has made one point definitively clear: Alger Hiss was guilty of spying for the Soviet Union.

Basics

In 1948 Alger Hiss was a member of America's foreign policy elite. Though the former assistant secretary of state had retired from government service, he landed a position as president of the Carnegie Endowment for International Peace. In stark contrast to the handsome, polished, and well-connected Hiss was his rumpled and diffident antagonist Whittaker Chambers.

In August 1948 Chambers publicly identified Hiss as one of several secret communists who had worked in the Roosevelt administration. Chambers's testimony to the House Un-American Activities Committee (HUAC) hit Washington like a bombshell. When Hiss denied all charges, the ensuing congressional hearings were the first such proceedings ever televised. Though the hearings resolved little, the most vocal interrogator, Richard Nixon, used the proceedings to impeach Hiss's character and launch his national political career.

Forced to defend his name, Hiss sued Chambers for libel. Pitting an obscure *Time* magazine editor versus Hiss seemed like a mismatch on paper. Rather than fold, Chambers doubled-down. Prior to facing a libel suit, Chambers had merely called Hiss a closeted communist. Forced to prove his accusations, he produced reams of classified documents that substantially raised the stakes.

263

Claiming the documents originated with Hiss, Chambers implicated both in a Depression-era Soviet espionage ring.

Though the statute of limitations for spying had expired, the evidence prompted a swift federal indictment for perjury against Hiss. Adding to the swirling intrigue were twists and turns worthy of a spy novel. Chambers had stashed much of the State Department's papers, including key 65 typewritten pages of copied classified documents, in a relative's dumbwaiter. In addition, he hid rolls of undeveloped film in a hollowed-out pumpkin. Though most documents were never hidden in vegetables, the press dubbed all the evidence the Pumpkin Papers.

Following the first trial, which ended with a deadlocked jury, a second trial began in November 1949. The proceedings hinged on two issues: the Pumpkin Papers and the Hiss-Chambers relationship. During the trial, it became apparent (even to Hiss's lawyers) the two were closer than the defendant admitted. Chambers (along with his wife and newborn son) not only stayed in Hiss's home for days and weeks at a time, but he also sublet an apartment, received a car, and borrowed money from Hiss. Despite the substantial proof of their friendship, Hiss only admitted to a cursory relationship with Chambers, whom he said used the alias George Crosley.

To convict, the prosecution had to establish Hiss as the Pumpkin Papers' source. Since hundreds, if not thousands, of officials had access to these documents, proving the source was difficult. At this point, the 65 typewritten pages became crucial. Experts eventually matched the documents' typeface to Hiss's old Woodstock typewriter. The typewriter, which Hiss had given to his maid years prior, was miraculously found in a garbage dump; a scenario adding to the case's oddities. On the strength of this and other circumstantial evidence, a jury found Hiss guilty of perjury. In the court of public opinion, however, Hiss was convicted of espionage.

Alger Hiss: The Roots of a Romantic Antifascist

Cultural baggage and the Great Depression's unique historical context have always limited the public's ability to accept Hiss's guilt. From James Bond to John le Carré spy novels, popular culture has cultivated a spy "archetype" that Alger Hiss fails to fit. Rather than asking for his drink "shaken, not stirred" and wooing women, Hiss was a nerdy bureaucrat. Despite Hollywood's archetype, in reality the yeoman's work of espionage is done without fancy gadgets, sexy sidekicks, or briefcases of cash. For every Aldrich Ames or Robert Hanssen, throughout the Cold War there were thousands of Alger Hisses, on both sides of the Iron Curtain, who spied for ideological beliefs.

Hiss's background helps historians understand his evolution from an ambitious self-made man into a romantic antifascist who risked everything to spy for the Soviets. Highly educated, extraordinarily well connected, and upwardly mobile,

Hiss lacked a radical family background, a taste for high living, or a penchant for self-destruction. He might have been raised on the East Coast and educated at Johns Hopkins University and Harvard Law School, but he was a self-made man who grew up in what one writer generously called "shabby gentility."

Though he was the second youngest of five, once his older sister and brother had died, the burden of advancing the family's fortunes fell upon Alger. From all appearances, he excelled under these heavy expectations. Exuding an aura of distinction and entitlement, defying his upbringing, one classmate said of Hiss, "If he were standing at the bar with the British ambassador and you were told to give a package to the ambassador's valet, you would give it to the ambassador before you gave it to Alger" (White 2004: 9).

Upon graduation Hiss landed a plum assignment as a legal secretary for Oliver Wendell Holmes. Though the secretaryship called for more social companionship than legal theorizing, the association with the acclaimed Supreme Court Justice almost guaranteed Hiss a bright future. Driven by his family's flagging fortunes and inner fortitude, Hiss was successful quite early in life.

Rather than buttress an understanding of how and why he became a spy, Hiss's background seemingly makes his later choices all the more puzzling. From a purely rational perspective, Hiss's choice to risk an intellectually and socially enriching career to engage in espionage confounds. The fundamental answer to the query is found in the extraordinary times during which Hiss came of age.

The Great Depression fundamentally transformed the reigning intellectual, social, and political paradigms of Hiss's day. With the American banking system on the brink of failure, unemployment hovering at 25 percent, and conventional wisdom utterly exhausted, young and ambitious idealists, like Alger Hiss, looked for alternatives. In this way, Hiss was not so different from millions of other liberals and leftists who drifted toward and flirted with radical political and economic reforms.

Hiss came to Washington at the Depression's height, when many believed centrally planned economies, like the Soviet Union, represented the future. In the early 1930s Joseph Stalin's brutality was mostly unknown and the Great Terror still lay in the future. Thus, during the early 1930s many, including Hiss, regarded the Soviet Union as a social democratic model embarking upon a brave and hopeful revolutionary path.

In tandem with a Depression-inspired zeal for radical political experimentation were the looming specters of fascism and another European war. Fascism's sudden rise and spread across Europe in the early 1930s pushed Western leftists toward the Soviets. Outraged by American isolationism and Britain's and France's seeming intransigence, leftists desperately sought to aid the antifascist cause. While some volunteered to fight the fascists in the Spanish Civil War, a few, like Hiss, battled the nazis by passing government secrets to the only power openly opposing fascism, the Soviet Union.

Prior to the nazis' rise to power, the Soviets had termed the Western noncommunist left "social fascists." Realizing Hitler had targeted the Soviet Union for extinction, the Kremlin capitalized on the nascent antifascist movement. Communist operatives where directed to cultivate and transform antifascist organizations into a "Popular Front," where communists and the noncommunist left joined together to defeat fascism (Weinstein 1997: 198). The Popular Front, however, remained a tactical rather than an ideological move. In the words of Comintern Chief Georgi Dimitrov, the Kremlin always intended to use the coalitions as "Trojan Horses . . . [to] penetrate the very heart of the enemy's camp" (Powers 1998: 121).

While the Popular Front officially came into being at the 1935 Communist International Seventh World Congress, prior to that year, the antifascist left was already hard at work. Indeed, Hiss's pro bono legal work for the International Juridical Association brought him into direct contact with a Popular Front organization and active communist party members. Thus, Hiss not only became more aware of what he called the "decrepit social structures" responsible for the Great Depression, he was already associating among some of the very communists he later conspired with before he arrived in Washington (White 2004: 28).

Like so many other young men of his educational background and elite connections, Hiss went to work for the New Deal upon Franklin Roosevelt's election. As a lawyer in the Agricultural Adjustment Agency (AAA), he was present for some of the New Deal's most hotly contested ideological debates. In this way, the AAA quickly divided into two camps: moderates and radicals. While the moderates wanted to solve rural America woes by helping landowners, the radicals (which included Hiss) believed sharecroppers and tenant farmers should receive top priority. Once it became clear that Secretary of Agriculture Henry Wallace would side with the moderates, the radicals looked for alternatives.

Frustrated, Hiss and his fellow AAA radicals gravitated toward the charismatic agriculture expert Hal
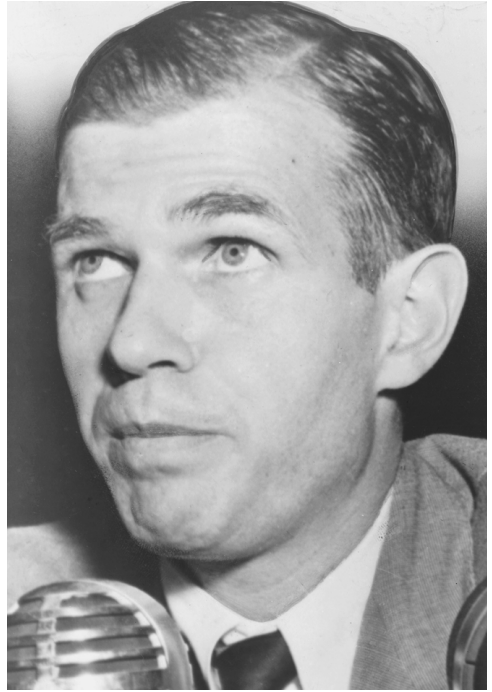
Alger Hiss testifies at his second perjury trial in 1950. Hiss, a former State Department official, was accused of lying during House Un-American Activities Committee hearings, during which *Time* magazine editor Whittaker Chambers identified Hiss as a member of the Communist Party. (Library of Congress)

Ware. Looking more like a Midwestern farmer than revolutionary, Ware was proof positive that appearances can deceive. Despite his all-American looks, Ware's mother was the "grand old lady" of the Communist Party of the United States, and Ware's innovative work in Soviet agriculture had earned Lenin's personal praise (Abt 1993: 39). Much more than a Jeffersonian farmer-intellectual, Ware was a highly trained Soviet operative who recruited Hiss to join a cell of communist sympathizers who worked in the Agriculture Department.

The Ware Group and Whittaker Chambers

The Ware Group, as it was dubbed, began innocently enough. Hiss's cell, like any other, met in secret, but in the beginning it was more coffee klatch than uber-spy network. Though members learned rudimentary "conspiratorial techniques," cell members mostly read and discussed communist texts or agriculture policy rather than pass state secrets to the Kremlin. Because Ware's reach was limited to the Agriculture Department, he was really grooming prospects for the time when one, or more, worked for the State, War, or Treasury Department (Tanenhaus 1998: 104–5).

In early 1935, Hiss left the Department of Agriculture to serve as chief counsel for the Senate's Nye Subcommittee. This post eventually led him to a position at the State Department. Hiss's promotion prompted the Soviet's Chief Intelligence Unit (the GRU) to separate him permanently from the Ware Group. As the center of a new and special "parallel apparatus," Hiss also received a new "handler," Whittaker Chambers.

By all appearances Chambers and Hiss made for an odd pairing. Obese, awkward, diffident, and brooding, Chambers was the seeming yin to Hiss's yang. The two did share humble upbringings, an Ivy League education, and a zest for ideas and literature. Moreover, because the GRU protected their star spy, Hiss now worked in "isolation." Rather than a community of like-minded communist intellectuals who shared the emotional struggles of leading double lives, as in the Ware Group, Hiss's only "comrade" was Chambers.

Defying protocol that called for limited contact to minimize opportunities for exposure, Chambers and Hiss socialized often. In addition to their friendship, Hiss's sheer productivity brought the two into more personal contact than normal. Unlike others who gathered documents for Chambers on a weekly basis, Hiss brought documents home every night. Rather than risk daily visits, the GRU had Hiss copy documents by typewriter (ironically, the source of the key 65 typewritten pages), which Chambers later couriered to the Soviets.

Chambers's Defection

In 1937 Chambers grew so fearful of an impending Soviet execution that he "defected." Chambers's worries were well founded. During the Great Terror, Stalin had ordered the arrest and execution of millions. Unlike the "famine" of

the early 1930s, during which Stalin had starved or executed farmers, the Great Terror directly targeted Communist Party officials, Red Army officers, and not a few spies. After several American-born Soviet spies were recalled to Moscow and reportedly killed, Chambers ignored his own travel orders.

Fearing for his life, Chambers defected. Before doing so, he visited his closest friend in the party, Alger Hiss. Failing to convince Hiss to defect, Chambers went underground and did not see his friend again until the HUAC hearings. Realizing that other spies-cum-defectors had been murdered, Chambers took a "life preserver," the infamous Pumpkin Papers, as insurance.

In 1939 Chambers resurfaced as a writer for *Time* magazine. At roughly the same juncture, the August 1939 Nazi-Soviet Pact prompted him to warn authorities about Soviet espionage in the State Department. Arranging a meeting with Roosevelt's aide, Adolph Berle, Chambers listed names of Soviet agents, which included Hiss. Though Berle took the information to the Federal Bureau of Investigation (FBI), it wasn't until 1945 that it was taken seriously. Lacking proof to press charges and seeking to avoid a very public controversy, Hiss was quietly forced to retire.

Allen Weinstein

From the time of Hiss's initial conviction through the mid-1970s, little new evidence surfaced. Despite the paucity of information, the case remained a hotly debated topic. In the years and decades following his conviction, Hiss maintained his innocence. In 1957 he penned a book, *In the Court of Public Opinion*, making that case. The zeitgeist of the late 1950s was not nearly as accommodating to Hiss's claims as were the 1970s. With one antagonist deceased (Chambers died in 1961) and another, Richard Nixon, essentially impeached for obstruction of justice, his chances for vindication increased exponentially.

Indeed, during the mid-1970s Hiss became a fixture on the university lecture circuit and in 1975 the Massachusetts Supreme Judicial Court even reinstated his law license. Indeed, with the key evidence coming from pumpkins, dumbwaiters, and garbage dumps, coupled with Nixon's track record for criminality, legitimate questions lingered.

Stepping in the fray was Allen Weinstein. A relatively unknown professor of history at Smith College, in 1973 Weinstein declared his intention to write a book on the case and prove Hiss's innocence. Claiming he would ignore the political fireworks and "confront the case itself," Hiss's proponents were delighted (White 2004: 146). With Weinstein filing a Freedom of Information suit for the release of FBI and Justice Department documents, Hiss gave him full access to his defense files.

While Weinstein began an exhaustive study of the case, Hiss launched an offensive to clear his name. In 1973 Hiss's son Tony interviewed his father for a

Rolling Stone piece, "I Call on Alger." The article preceded Tony's 1977 memoir of his dad, *Laughing Last* and John Chabot Smith's 1976 sympathetic work, *Alger Hiss: The True Story*. Though none of these pieces offered any new substantial evidence, they helped Hiss's case for exoneration gain momentum.

As Hiss moved toward absolution, Weinstein researched. Assuming the FBI files would reveal substantial flaws in the prosecutor's case, Weinstein pored over records to no avail. Next, he turned to what became his most definitive evidence: Hiss's own defense files. Surprising himself, Weinstein was shocked to find Hiss's own lawyers doubting their client's truthfulness.

Moreover, to Weinstein, the defense files revealed Chambers's truthfulness and Hiss's pattern of lies. As Weinstein wrote, "[a]t every important point Chambers's testimony about his underground experiences and . . . about Hiss's complicity, checked out" (White 2004: 177). For example, during the perjury trial, Hiss claimed the Woodstock typewriter had gone to a junk dealer. However, a letter in Hiss's own hand, written near the time of that testimony, revealed he remembered giving the typewriter to his maid. This is merely one example of many where Weinstein uncovered evidence of Hiss's lies.

In addition to the infamous typewriter was the Bokhara rug. The rug in question was not only a further item connecting the two, it also revealed yet another lie and an additional piece of evidence linking Hiss to espionage. Hiss had always claimed his contact with "George Crosley" had ceased in 1936, while Chambers said their friendship ended in 1938. To prove his timeline's veracity, Chambers recalled giving Hiss a Bokhara rug in 1937.

Hiss still possessed the rug but claimed Chambers gave it to him in 1935 as partial payment for rent owed on a sublet. Weinstein found the man who bought the rugs on Chambers's behalf, Meyer Schapiro, and later located the dealer who originally ordered and shipped the items. Moreover, Schapiro, a Columbia University art professor and former communist spy, admitted he ordered three rugs, which were all distributed as gifts to Soviet agents, one of whom was Hiss. With receipts and eyewitnesses attesting to the veracity of Chambers's story, Weinstein caught Hiss in yet another lie.

In addition to Schapiro's information regarding the rug, Weinstein found letters corroborating Chambers's story of the Pumpkin Papers' source. In letters between Chambers and radical journalist Herbert Solow, dated 1937 and 1938, Chambers specifically mentioned the "life preserver of stolen government documents" he had squirreled away as insurance against Soviet retaliation for his defection.

Though Schapiro and the Solow letters helped solidify Chambers's claims, 15,000 pages of declassified FBI documents provided the final proof of Hiss's guilt. Proponents of Hiss's innocence had long maintained that the FBI had framed Hiss and were positively gleeful when American Civil Liberties Union (ACLU) lawyers convinced the federal courts to order the documents' release.

Weinstein, however, found no evidence of an FBI conspiracy to frame Hiss. Though the bureau had been inept at times, the documents only served to buttress Chambers's version of the events.

Weinstein's 1978 book *Perjury: The Hiss-Chambers Case* painstakingly detailed the evidence of Hiss's guilt. While the work enjoyed critical and popular success, Hiss and his defenders successfully cast enough doubt on the author's research to maintain doubt in the public's eyes.

By 1980 a very favorable documentary, *The Trials of Alger Hiss*, was released to favorable reviews and audiences. Following the movie, in 1983, PBS's *American Playhouse* produced a dramatic reenactment of the Hiss case. Titled "Concealed Enemies," the television show depicted Hiss as an idealist who fell prey to McCarthyism and overly zealous anticommunists. Three decades after Hiss's perjury conviction, Americans remained divided on the case.

The Post–Cold War World: American and Soviet Archives

Until the late 1990s, it was intellectually defensible to question Hiss's guilt. Though conventional wisdom regarded Weinstein's book as definitive, many high-profile scholars and public intellectuals supported Hiss. Whether one challenged Whittaker Chambers's truthfulness, Richard Nixon's prosecutorial scruples, or the reliability of evidence literally stored in a pumpkin or dumbwaiter, legitimate questions lingered. Today, however, evidence from Russian and American archives proves Hiss's guilt beyond a reasonable doubt.

In the early post–Cold War era, a brief window of opportunity gave researchers unprecedented access to Russian and American intelligence archives. From the U.S. army's Venona Project to the KGB's "Operational Correspondence Files," these new sources definitively reveal Hiss's guilt. According to these sources, from 1934 through 1945 Hiss passed thousands of classified documents to the Soviet Union, actions that unquestionably damaged U.S. national security and substantially shaped the early Cold War.

Hiss's supporters were temporarily elated when a former official historian for the Soviet government, Dmitri Volkogonov, proclaimed Hiss's innocence. Though the *New York Times* editors said Volkogonov "would not lightly render an official opinion without being sure of his research," they were wrong. Within weeks Volkogonov admitted he had spent two days in the archives researching the case and retracted his sweeping pronouncement.

The entire episode provided sufficient impetus for the Russian and U.S. governments to give professional historians access to the proper archives. Unlike traditional social, diplomatic, or political history, however, the history of espionage calls for a different sort of research methodology and expectations. Indeed, trying to decode historical fact from letters or memos is hard enough. Deducing fiction from fact in communiqués designed to hide identities and

actions is even more complicated. Thus, rather than search for a definitive document completely exonerating or convicting Hiss, historians hoped the archives would add pieces of evidence to what was already known.

Two leading experts on U.S. domestic communism, Harvey Klehr and John Earl Haynes, first combed the Russian archives. Ironically, it was in a foreign archive where they learned about an American espionage program, the Venona Project. A wartime venture of the U.S. army's Signal Intelligence Service, Venona employed linguists, mathematicians, and cryptanalysts to decipher Soviet cables. Four years of intense labor led to a breakthrough that enabled U.S. intelligence officials to decipher Soviet diplomatic cables from 1942 until 1946 when a Soviet spy alerted the Kremlin.

The deciphered cables silently rocked government leaders. Though officials had hoped Venona would offer insight into Soviet diplomacy, analysts learned much about Soviet espionage in the United States. In addition to the 349 U.S. citizens, permanent residents, and immigrants actively spying for the Soviets, the cables revealed evidence of high-level government officials spying for the Soviets. From the Treasury Department's second ranking official Harry White and Franklin Roosevelt's personal assistant Lauchlin Currie to nuclear physicist Klaus Fuchs employed in the Manhattan Project, Venona uncovered a deep web of Soviet espionage.

Venona also implicated Alger Hiss, which in addition to the wartime testimony of Chambers and another spy-cum-defector Elizabeth Bentley finally led to his quiet dismissal from the State Department. The Venona cables were never revealed because U.S. officials refused to offer any public information about a top secret program. Indeed, U.S. officials used what they learned from Venona to try and break Soviet diplomatic code throughout the Cold War.

Though the cables were deciphered, most communiqués still used code names. Only after piecing together a series of cables that offered tidbits of personal information could analysts make an educated guess connecting specific individuals to code names. Luckily one of the most specific Venona cables offered identifying characteristics for an American agent, code named "Ales." According to Cable 1822, "Ales" had begun spying in 1935, led a small group of spies that included his relatives, was present at the Yalta Conference, and attended the following Moscow Conference.

While the first three characteristics clearly applied to Hiss, the traits could have just as easily fit hundreds of other State Department officials. The most damning evidence in Cable 1822 was Ales's presence at the Yalta and Moscow conferences. Indeed, Hiss was one of only four State Department officials who attended the Moscow Conference following the Yalta gathering. Among those four officials, none, except for Hiss, had ever had any ties or allegations of espionage.

It was not merely Klehr and Haynes who mined the new archives but Weinstein as well. With unprecedented access to the KGB's Operational Files, Weinstein and his coauthor Alexander Vassiliev uncovered new information that

---

Nathaniel Weyl, Alger Hiss's Accuser

Alger Hiss was accused of being a member of the Ware Group in 1933—the Ware Group advocating that the U.S. government adopt procommunist policies. These claims were made before the McCarran Committee by Nathaniel Weyl (1910–2005), an economist who had been a member of the Communist Party of the United States.

Weyl had been born in New York City, his father being a prominent progressive writer who had helped found the *New Republic*. The young Nathaniel Weyl went to Columbia University and then to the London School of Economics, becoming interested in communism and joining the Ware Group. Totally disenchanted following the signing of the Molotov-Ribbentrop Pact of 1939, he went to work for the Federal Reserve Board and served in the U.S. army for two years. He later became a dedicated anticommunist, studying the role of communism in Latin America, being sharply critical of Fidel Castro and the Cuban Revolution.

---

corroborated the veracity of the Venona cables. The KGB files revealed and reconfirmed that hundreds of U.S. agents, including Hiss, labored for the Kremlin. Thus, it was not merely Chambers's testimony, Hiss's defense files, or even the Venona cables that confirmed Hiss's guilt, but the KGB archives as well.

The Vassiliev Notebook

Despite the overwhelming evidence, Hiss's defenders continued to mount a substantial posthumous defense after Hiss's 1996 death. They not only aimed their fire at Weinstein but also at his Russian researcher and coauthor Alexander Vassiliev. A former operative in the KGB Intelligence Directorate's "American Department," who left the agency for a career in journalism, Vassilev spent two years researching the KGB Operational Correspondence Files. These papers detailed the correspondence between spies and their handlers and the handlers and their KGB bosses. The archives also offered personal files on each agent.

As someone who understood the KGB's culture and the world of espionage but possessed a journalist's nose for investigating a story, Vassiliev was ideal for the task. During the mid-1990s, when economic unrest and post–Cold War malaise rocked Russia, authorities were pleased to have Vassiliev coauthor books detailing the KGB's prodigious successes in espionage. Though Vassiliev was telling a story the Russian defense establishment wanted, they did not allow document photocopies. Instead, they let Vassiliev use a notebook to take painstaking notes of the documents.

Because U.S.-Soviet relations have cooled since the mid-1990s, scholarly access to KGB archives has ceased. Thus, Vassiliev's 1,115 pages of notes are the most comprehensive source on Soviet espionage available to the public.

Beyond his notebook Vassiliev is a unique source in and of himself. Separated by decades and a continent from the *sturm* and *drang* of the Hiss case, Vassiliev has no ideological ax to grind. Before he cowrote a book with Allen Weinstein, he had barely even heard of Alger Hiss. Indeed, he claims, "I don't give a damn about Alger Hiss. Never did" (quoted from Haynes and Klehr 2009: liii).

Vassiliev might not have cared about Hiss, but what he uncovered brought the ire of his advocates. Because the KGB files were poorly organized, security officials could not have restricted access to materials even if they had wanted to. Consequently, even though Russian intelligence protocol is to keep a spy's identity secret at all costs, due to the archive's muddled nature, Vassiliev found a document written by the Soviet's Washington, D.C., station chief that listed the names and code names of U.S. agents. The "Rosetta Stone" of Soviet espionage reveals "Ales" was indeed Alger Hiss.

Conclusion

More than any single American, Alger Hiss is singularly associated with the rise of Senator Joseph McCarthy. To critics, McCarthyism and anticommunism became synonymous. An inveterate liar and drunkard, Joseph McCarthy used the Hiss case and communist subversion as a partisan cudgel to save his own sinking political career and smear political opponents. In the early to mid-1950s, McCarthy ruined thousands of lives and careers with his reckless investigations and unfounded charges.

For those wishing to cast aspersions on the Cold War in general and anticommunism in particular, McCarthy has become a convenient foil to paint the entire enterprise with the same broad brush. Whatever one thinks about the Cold War, McCarthy, or Vietnam, it is important to separate one from the other. To endorse the basics of the Cold War and anticommunism is not tantamount to approving of McCarthy and Vietnam.

Since the Soviet Union's founding, the U.S. right and left had been bitterly divided over a response. In the spring of 1919, the Comintern's call for revolution led to a spate of domestic bombings, gaining American's attention. With the FBI and attorney general jailing foreign-born radicals without a shred of evidence, the Red Scare and conservative anticommunism were born. Defined by nativism and a disregard for unconstitutional procedures, the conservative anticommunism of the 1919–1920 Red Scare returned in the form of Joseph McCarthy.

Similarly, the Red Scare also spawned the left's response. Terming anticommunism a "false and dangerous attack on . . . civil liberties" that used communism as a guise to assault the Constitution, the ACLU established the left's response (Powers 1998: 33).

Though the nascent Soviet Union hardly represented an existential threat, as the Alger Hiss case reveals, the Kremlin's intentions were hardly benign.

Indeed, Eastern European, Chinese, and Russian archives have revealed Joseph Stalin's aggressive intent. Despite what some revisionists claim, the Cold War was not merely a conflict between morally equal superpowers fighting for hegemony. Rather, as the dean of Cold War scholarship John Gaddis writes, "Stalin's was, therefore a grand vision: the peacefully accomplished but historically determined domination of Europe" (2006: 14).

Mirroring Gaddis is Melvyn Leffler who claims the early Cold War was defined by Joseph Stalin and his nearly limitless fear of "capitalist encirclement" (2007: 21). Communist ideology told Stalin that another capitalist war was historically predetermined. Resolved to avoid another invasion, Stalin imposed communist regimes across Eastern Europe (and sought to do so in Western Europe). In response, Harry Truman opted for the Marshall Plan, Truman Doctrine, and containment. Seeing Truman's policy as capitalist encirclement, Stalin supported revolutionary nationalist movements as the route to toppling the capitalist world.

With the Kremlin committed to fomenting revolution and the United States opting for containment, the Cold War was born. For a brief period, however, Truman, liberals, and key conservatives forged a bipartisan foreign policy based on reasonable anticommunism. Unfortunately, the Alger Hiss case drove a deep and lasting wedge into that consensus. Once McCarthy emerged, the ACLU's repudiation of all anticommunism gained credence with the left and a generational divide was born (or reborn).

Despite McCarthy's blathering and the left's willing disbelief, during the Great Depression and World War II, Soviet espionage reached into the State Department and even the Oval Office (in the form of FDR's personal assistant Lauchlin Currie). A few agents produced significant information. For instance, agents passed the technology and research data for the atomic bomb and a plethora of other military technologies. This enabled the Soviets to produce an atomic bomb, years earlier than otherwise possible. According to John Earl Haynes and Harvey Klehr, the Soviet advent of the atomic bomb emboldened Stalin and led directly to the North Korean invasion of the South (Haynes and Klehr 2009: 545).

Ironically, once McCarthy began his campaign, the golden age of Soviet espionage was over. Through the decades, the Kremlin recruited spies in the Central Intelligence Agency (CIA) and the State Department, but they were not part of a wider ideological movement. Alger Hiss was a spy. Even more damning, he spent decades lying to friends, family, and the larger intellectual community, which fomented a larger cultural and ideological war over Vietnam and the Cold War. Indeed, Alger Hiss's innocence eventually became an issue of faith among some on the left. In this way, the entire sorry episode reveals the lures and dangers of making political beliefs into a secular religion.

Alger Hiss was never representative of anything larger than one man's ideological path through an extraordinarily complicated time. He was a spy who

betrayed his country while aiding a regime that murdered tens of millions. There is no object lesson in that.

References and Further Reading

Abt, John. *Advocate and Activist: Memoirs of an American Communist Lawyer*. Urbana: University of Illinois Press, 1993.

Gaddis, John. *The Cold War: A New History*. New York: Penguin, 2006.

Haynes, John Earl, and Harvey Klehr. *Spies: The Rise and Fall of the KBG in America*. New Haven, CT: Yale University Press, 2009.

Hiss, Alger. *In the Court of Public Opinion*. New York: HarperCollins, 1957.

Leffler, Melvyn. *For the Soul of Mankind: The United States, the Soviet Union, and the Cold War*. New York: Hill and Wang, 2007.

Powers, Richard Gid. *Not Without Honor: The History of American Anti-Communism*. New Haven, CT: Yale University Press, 1998.

Tanenhaus, Sam. *Whittaker Chambers: A Biography*. New York: Random House, 1998.

Weinstein, Allen. *Perjury: The Hiss–Chambers Case*. New York: Random House, 1997.

White, Edward. *Alger Hiss's Looking Glass Wars: The Covert Life of a Spy*. Oxford: Oxford University Press, 2004.

CON

Alger Hiss was a man of principle and conviction. He dedicated himself to public service, and his repayment for that was a life and name tarnished by a vigilante crusade against communist infiltration in the U.S. government. His trial and conviction were little more than an attempt to feed and fuel a media frenzy, maintain a "public health campaign," and give justification for the use of tax payer dollars to build bomb shelters, weapon arsenals, and encourage the average citizen to do the same in his own backyard.

Hiss's upbringing, framed by tragedy and hard lessons, provided him with the strong foundations he needed for a life in public service. These core values are what steered him through defenses of President Franklin Delano Roosevelt's New Deal programs, through the formation of the United Nations during World War II, and through two national trials. Even though he was convicted, he fought to restore his name and membership in the Massachusetts State Bar Association, and he never denounced his country through flight. The life of Alger Hiss and his legion of friends and mentors who hold prominence in U.S. history show that he could not have possibly been a spy. Alger Hiss took the blame, in a moment of

national hysteria, for actions that he could not have committed. He lived his life so vividly in the public spotlight that secret meetings and phone calls would have been impossible to complete, and his moral center would have prevented them. More so, the evidence continually shows that he did not betray his country.

Early Life and Education

Alger Hiss was born in November 1904, the fourth of five children, to Mary Lavina Hughes and Charles Alger Hiss. Hiss and his four siblings grew up in Baltimore, Maryland, where his father was a middle-class grocer. This early life contrasted the upbringing of his counterparts in Washington, D.C., politics, as Hiss did not have access to private schools, wealth, and luxury. In many ways it set the stage for the tragedy and turmoil he would endure later on in life. When he was two his father committed suicide, leaving his mother to struggle to support the family. They lived in what was dubbed a neighborhood of "shabby gentility," and his mother often relied on the kindness of family members and friends to help support the family. The Hiss children did not fair well—when Alger was 22 his older brother Bosley died of Bright's disease, and at 25 he lost his sister Mary to suicide. Yet, in spite of these setbacks this young man—who would later be called well bred—triumphed over personal tragedy.

He earned his undergraduate degree from Johns Hopkins University with Phi Beta Kappa honors, and his classmates voted him most popular for his charm and wit. This bright and ambitious student continued his education at Harvard Law School from which he graduated in 1929. While at Harvard Law, Hiss studied under Felix Frankfurter. Frankfurter would later be appointed to the U.S. Supreme Court under President Franklin Delano Roosevelt in 1939. Frankfurter recommended Hiss as the private secretary to Supreme Court Justice Oliver Wendell Holmes, for whom he served a year. Hiss then worked for a Massachusetts law firm and then a New York law firm until 1933, when his former mentor sent him a telegram saying his service was needed at the national level. Hiss took the words of Frankfurter to heart, and he entered the Washington, D.C., arena that year. He worked directly with FDR's New Deal team, and Hiss specifically undertook the battles of the Agricultural Adjustment Administration (AAA). His work with displaced farmers—who had either lost their farms, lacked farm income, or had crop failures—reinforces Hiss's character and upbringing. He started his public service career fighting to protect and defend the nation's farmer, grocer, and family.

Rising of a Public Service Career

After serving with the AAA he worked with the Nye Committee, which devised the "Merchants of Death Thesis," pointing the finger for U.S. involvement in

World War I to big business and greedy industries. Throughout his endeavors with FDR's New Deal reforms, his skills were used to defend the constitutionality of the governmental programs and justify the initiatives as benefiting the people. Much of Hiss's career would rest on him working for the benefit of the larger good, justify the actions of the federal government, and helping the American people understand how a stronger and bigger arm of governmental intervention would benefit their daily lives and not hinder their personal growth and freedoms.

In 1936 Hiss, along with his younger brother Donald, went to work for the State Department. As Hiss worked his way through the State Department he advised on Far Eastern Affairs, and most notably he served as deputy director of Special Political Affairs, which led him to directly engage in the construction of the United Nations Charter. In 1945 Hiss—a member of the Yalta Conference—led the opposition to Stalin's proposal for 16 Soviet votes in the UN General Assembly (Stalin received three). Hiss served at the center of the creation of the United Nations, when in 1945 he presided at the San Francisco Conference as secretary general. Hiss often remarked that the highlight of his public service career came when he was asked to hand-carry the UN Charter to Washington, D.C., for President Truman's signature.

Finally, after this very public and international career, centered on the preservation of community and national vitality, Hiss left public service in 1946. Even with his change of venue his grassroots threads did not leave him, as he took on the role of president of the Carnegie Endowment for International Peace until May 1949. Then, with abrupt ferocity, Hiss's life and career came under national and international scrutiny, becoming the center of debate and analysis, and he became an object of national hatred.

Politics of the Moment

In 1948 Alger Hiss was formally charged with two counts of perjury by a federal grand jury. These charges arose during testimony before the House Un-American Activities Committee (HUAC), which formed in 1934 to investigate and counter nazi propaganda. Quickly the committee changed its goals, as World War II ended. The ending of the war forced the United States, and other Allied powers, to reexamine themselves in light the Soviet Union's communist government and Stalin's avowal to expand his communist influence throughout Europe, and the complexity of the friendships during the war had to be reexamined. The HUAC dealt with much of the communist hysteria during the 1950s, and trials about comic book content even appeared before them in the early 1950s.

With the ending of World War II, the United States found itself on the defensive against its former ally. In what set the stage for the Cold War and a progressive arms race, in 1945 the HUAC became a permanent committee of

Congress. As early as 1947, federal government employees had to take loyalty oaths and denounce ties to communism, and the same year saw nine sensational days of testimony about the possible threat of communists overtaking the motion picture industry. Clearly, the nation was on edge, and the widely broadcasted activities of the HUAC did little to ease individual concerns. Instead, local and national media campaigns instructed individuals on how to spot a communist and detailed how communism would destroy the family, community, and "American way of life." Films and editorials defining commies, and communistic activities, routinely relied on vague descriptions of communist behaviors and provided few (if any) concrete examples of explicitly communist activities. Most commonly, persons who regularly spoke out against the United States, stood for organized labor, socialist beliefs, denounced imperialism, or even supported equal rights were callously labeled as communists.

More potent forms of propaganda, and scare tactics, told people to build "backyard bomb shelters" and stock up on supplies in case Communist Russia decided to attack the United States. These posters, radio announcements, and film shorts were made and endorsed by the federal government, which inadvertently gave them credibility in the public's eyes. And, of course, the nine days of nerve-wracking testimony from the HUAC passionately declared that communists had infiltrated Hollywood. On November 25, 1947, 10 writers and directors found their names blacklisted; they became known as "The Hollywood Ten," and the list was dubbed the Waldorf Statement (released at the Waldorf Hotel in New York City). The newly formed Motion Picture Association of America sanctioned the list, as if to give it a seal of approval and moral authority, and later the Motion Picture Association gave its backing to pamphlets like *Red Channels*. *Red Channels* named 151 Hollywood names, and these activities led to the blacklisting of almost 300 Hollywood artists. Many of these artists would never work in the industry, several fled U.S. borders and worked in Mexico and Latin America, and future U.S. President Ronald Reagan even testified before HUAC naming names. Clearly the nation was in the midst of an emotional crisis. These actions, and a plethora of others, set the stage for the naming and false conviction of Alger Hiss.

The Trial and Evidence

On August 3, 1948, Whitaker Chambers, senior editor of *Time* and a former communist, told the HUAC that Alger Hiss worked for underground Communist Party organizations (the Ware Group). At this point, Chambers made no mention of espionage, later he would change his story, and on several occasions he was forced to admit that he had lied on the stand. Even though Chambers gave shaky testimony, Hiss already had a case being built against him. In 1942 the Federal Bureau of Investigation (FBI) actively began to keep a file on Hiss and his so-called espionage activities. A further collaboration of Chambers's weak

Two guards flank Elizabeth Bentley, a former American spy for the Soviet Union, at a meeting of the House Un-American Activities Committee where Alger Hiss (extreme right), a former State Department official, denied charges that he was a member of the Communist Party. (Library of Congress)

testimony weighed on the witness of Elizabeth Bently, an American spy for the Soviet Union. In 1945 she defected and told the State Department that a "Eugene Hiss" had been in contact with Soviet officials, and Igor Gouzenko defected to Canada the same year telling a tales of an unnamed assistant to the secretary of state.

In an earnest effort to save his reputation Hiss volunteered to testify before the HUAC on August 5. Richard Nixon, an aspiring politician, pressured the committee to continue the investigation, even though there appeared to be little (if any) hard, substantial evidence. Nixon used the writings of John Francis Cronin, a Roman Catholic priest who had gained access to FBI files, as justification for the continued investigation. The investigation culminated in having Hiss identify Chambers from a photograph, and he then was requested to meet with Chambers in a hotel suite with HUAC members present. Hiss declared that he knew Chambers as "George Crosley." Hiss attested that Chambers/Crosley had claimed to be a freelance writer, he had sublet him (Chambers) an apartment a decade earlier, and he had given him an old car.

Enraged, Hiss invited Chambers to make his charges against him in a public setting (since congressional testimony remained protected from libel charges).

Typewriters and the Hiss Case

Martin Tytell was the typewriter expert who gave evidence on behalf of Alger Hiss at his trial, and whose statements were particularly important in proving that Hiss was not guilty. He was born on December 20, 1913, in New York, his parents being Russian Jewish immigrants. He attended a local school and then went to Saint John's University and then New York University.

In 1938 Tytell established the Tytell Typewriter Company and managed to develop typewriters that could be sold for far less than many of his business rivals. He also managed to design a coin-operated typewriter whereby people could use the machine for 10 cents per 30 minutes.

During World War II, he served in the U.S. marine corps but was important in creating foreign language typewriters that were used by the military. Indeed after the war his company boasted that they could make typewriters that could be used for 145 languages, including designing carriages that worked in reverse for typing messages in Hebrew or Arabic. He died on September 11, 2008, after a long illness.

Chambers called Hiss a spy on *Meet the Press*, a popular radio show (which in the modern era became a Sunday morning politics television program). Hiss began his libel lawsuit, but on November 17, 1948, Chambers produced 65 pages of photocopied government documents "proving" Hiss was a spy (the "Baltimore Documents"). On December 2 he led investigators to five rolls of 35 mm film he had hidden in hollowed out pumpkins at his family farm in Maryland. These became known as the Pumpkin Papers. The story behind the documents said that Priscilla Hiss—Hiss's wife—had retyped State Department documents on the family typewriter, creating the Baltimore Documents, and of the five rolls of film only two had State Department documents and one roll was blank as it had been severely overexposed.

The Charges

Based on the flimsy evidence in the Baltimore Documents and Pumpkin Papers, Hiss was charged with two counts of perjury, and Chambers never faced a jury or charges. Hiss did not receive espionage charges because the statute of limitations had expired. At trial, along with the so-called documents provided by Chambers, the defense located the old family typewriter (given away years earlier) and presented it as evidence. The Woodstock had been given to a family servant, and Hiss's council produced the machine when they learned the prosecution was close to discovering it. Expert testimony could only say that the documents had probably been typed on the typewriter, and Hiss's camp maintained the argument that the documents had been forged. The first trial ended with a hung jury, and the second convicted Hiss in January 1950. On January

25, 1950, Hiss received a sentence of five years, which he served 44 months. Hede Massing, an ex-communist, testified at the second trial, providing a weak backing to the charges. The added testimony of Massing, circumstantial as it was, seemed to confirm that Hiss had seen Chambers after the last 1937 encounter that he affirmed. He was then convicted on two counts of perjury, and not espionage, which has fueled the fire to the controversy.

At the second trial, defense lawyers brought in experts to analyze the mental state of Chambers, since he had already admitted that he had lied on the stand. Hiss's attorneys failed to paint Chambers as unstable when they bungled their attempt to criticize Chambers's translation of Franz Werfel's *Class Reunion* into English. The prosecution asked what Chambers's translations could show about his personality, as he had also translated *Bambi*. The headlines roared "Who's psycho now?" Both sides were clearly trying to grasp on to every thread of hope for their case. These outlandish tactics by the lawyers show just how far flung and shabby much of the evidence against Hiss really was.

Conviction

Hiss's conviction reinforced the mindset that there was a commie around every corner, and his high-profile court loss gave justification to FBI searches, seizures, and infringements upon civil liberties. Most important, in a perverse manner, it gave the American public a sense of security knowing that commie infiltration had been thwarted (well at least for the time being). Two weeks after the Hiss conviction, on February 9, 1950, Senator Joseph McCarthy gave his now infamous Wheeling, West Virginia, speech in which he claimed to have the names of 57 communists in the federal government (that number would later jump to 205).

The shaky evidence aside, transcripts and other documents have been released in the 50-plus years since his unjust conviction. Hiss's libel suit fell short, as his perjury charges were brought forth, and during his 44-month prison term he served as a lawyer to inmates while fighting to reinstate his good name. Upon release he worked for a stationary company, he and Priscilla divorced in 1959 (he did not remarry until after her death in 1986), and in 1957 he published his book *In the Court of Public Opinion*. He argued that the Baltimore Papers had been forged, and he consistently claimed that he had been railroaded. Richard Nixon in *Six Crises* (1962) said that duplicating machine typefaces was not possible in 1949, but *The Quiet Canadian: The Secret Service Story of Sir William Stephenson* by H. Montgomery Hyde (1963) shows readers how this was done as early as 1936.

In accordance with the Freedom of Information Act, on July 31, 1975, Hiss won his suit against the Justice Department for the release of the Pumpkin Papers. The documents revealed an overexposed roll of film, two with documents introduced at the two trials, and two more with nonclassified State Department

documents about navy matters of life rafts and fire extinguishers. The data found on these papers proved complexity to uphold the argument that Alger Hiss committed acts of espionage with these materials. Hiss used this evidence as the basis for his claim of prosecutorial misconduct, which was denied. Then, in a truly remarkable turn of events, Hiss was readmitted to the Massachusetts Bar Association on August 5, 1975. He became the first person to be readmitted to the Massachusetts Bar after a criminal conviction. Clearly, this action points in the direction of Hiss's innocence, as the Massachusetts Supreme Court saw the flimsiness and weak charges on which he was convicted. The state's Supreme Court gave a unanimous ruling for Hiss on the grounds that he demonstrated moral fortitude and intellectual prowess for the profession. This same year the HUAC was disbanded as a standing congressional committee.

Yet, even though Hiss fought for his right to practice law again, he appeared on Howard K. Smith's *News and Comment* in 1962 for a segment on Richard Nixon. Nixon had just lost the California governor's race, and the show erroneously predicted his death from politics. The show was canceled a year later, after an onslaught of protests about the airing of a convicted perjurer. Hiss may have won the support of justices, but the public still held him at bay.

Recent evidence has continued to fuel the debate about Hiss's innocence. Aside from Nathaniel Weyl's 1952 testimony before the McCarren Committee saying that he had been a member (along with Alger Hiss) of the Ware Group in 1933, his book stated that he did not fully believe in Hiss's conviction. Weyl never testified at the Hiss trials, for reasons unknown.

In 1976 the Freedom of Information Act forced the Department of Justice to release FBI and prosecution records about the case, and in July 1978 Hiss's defense team asked for *coram nobis*. *Coram nobis* would have overturned the guilty verdict on the grounds of prosecutorial misconduct, but the petition was denied. The grounds Hiss used for this failed defense are that the FBI withheld evidence from Hiss that typewritten documents could be forged, that the FBI knew there were discrepancies with the serial number of the Woodstock recovered and entered into evidence, that Horace W. Schmahl worked as a detective for Hiss while reporting to the federal government, and that the FBI illegally monitored Hiss. These petitions were denied in 1982 and 1983.

Other points of contention with the evidence reveal that the Woodstock did not correlate to manufacturer records, via serial numbers, and a Nixon White House staff member John Dean claimed that Nixon told Charles Colson that the HUAC did fabricate the Woodstock. No substantial evidence has come to support this 1976 statement, but it does add flavor to the continual array of circumstantial and secondhand information that form this case. The legacy of Nixon's lying to federal authorities and his wire tapping in the Watergate Affair also shed a more sinister light on Hiss's conviction. More scholarly research

questions the fairness of Hiss's trials with Richard Nixon at the helm of the prosecution. The testimony of two former Soviet spies also adds fire to this mix of possible governmental tampering with evidence.

In 1992 Hungarian records showed that Noel Field named Hiss a communist spy while he was imprisoned from 1949 to 1954 in Hungary. Field claimed that in 1935 Alger Hiss tried to woo him to the Soviet side, but Field told him he was too late. This story is similar to the one that Hede Massing told U.S. authorities in 1947 after her defection. After Field's release he wrote a letter to the Communist Party's Central Committee in Moscow decrying many of his statements in prison. He said that the torture made him confess lies as truth, and Hiss used this letter as added support that the naming of him was in error.

In 1992, after the fall of the Soviet Union and the former communist bloc in Eastern Europe, Alger Hiss wrote to Russian historian Dimitry Antonovich Volkogonov. Volkogonov oversaw Soviet intelligence archives, and he pleaded with him to release the files connected to his case. On October 14, 1992, Volkogonov complied by releasing a report that attested no evidence connected Hiss to the Soviet intelligence units. Later controversy emerged when Volkogonov admitted that he only spent two days researching this issue, but memoirs of several Soviet officials clear Hiss's name. Finally in 2007 Svetlana Chervonnaya testified that Hiss's name remained absent from Soviet intelligence records. For a moment it looked like Hiss's name would finally be cleared, once again, lending fuel that the defected spies were spinning tales to win freedom and asylum.

In 1995 the existence of the Venona Project was made known to the public, and its connection to Hiss adds another layer of complexity to this tale of deceit and secrecy. The project decrypted Soviet intelligence messages from 1942 to 1945, and amid the mounds of suspicions another circumstantial revelation arose. FBI Special Agent Robert Lamphere said the Soviet spy called Ales was probably Hiss. Once again, unsubstantiated evidence arose to denounce Hiss, but scholars have continued to dispute it. Not only has the scholarly dispute rested on the earlier knowledge that the federal prosecutors gravely mishandled Hiss's earlier trial(s), but the foundations of the Ales identity did not correlate to Hiss. Ales was said to be a leader of espionage agents; Hiss was accused of acting alone. Ales was a GRU (Soviet's Chief Intelligence Unit) agent and obtained military intelligence; at his trial Hiss was only accused of releasing nonmilitary data. It remains unlikely that Alger Hiss could have been a spy after 1938 as Chambers broke from the Communist Party and went into hiding, and Hiss's accession in Washington, D.C., made him a highly politicized public figure. Keeping his identity secret after 1938 would have been near impossible and risky. Ales did continue his espionage activities. Finally, records place Ales in Mexico City when Hiss was clearly in Washington, D.C.

John Lowenthal, Hiss's lawyer, argues that the one Venona cable (1579), which names Hiss, could not be pointed directly at Alger. Instead, his brother

Donald could be the name, and on a more stifling note the GRU would have avoided naming Hiss at all. Using a spy's real name, instead of a code name, in any document remained unorthodox. In April 2007 scholars gave evidence to support U.S. Diplomat Wilder Foote as Ales, but as with so much of this case, the Foote hypothesis is also in dispute. The dispute rebuffing Foote as Ales rests on counter-arguments saying that the documents had been misinterpreted and that the Soviet cable was mistaken in saying that Ales was in Mexico City. Connecting Hiss to Ales stems from a 1990 book by Christopher M. Andrew and Oleg Gordievsky that names Hiss as Ales. Gordievsky's source for Ales was a journalist, and his account has also fallen under scrutiny. Nonetheless, the continual battery of evidence for and against Hiss, compounded with the grave misconduct by federal prosecutors, shows that the light of innocence still shines in Hiss's favor.

Alger Hiss died on November 15, 1996, still claiming his innocence.

Conclusion

The irony of the Hiss case is not only in the fact that he was a well-received public servant, mentored by some of the biggest names in politics and U.S. history, but also that he was a central character to the formation of the United Nations and its charter, and numerous scholars note that the United Nations and the Yalta Conference set the stage for the Cold War and postwar world. Hiss adamantly fought against the Soviet Union in the UN debates, and he continually worked for the progress of a better national center. The railroading hysteria of the immediate postwar years took away voices of reason, clarity, and balance in U.S. politics. Alger Hiss is only one character, in a cast of many, who figuratively took a bullet for his country without just cause.

References and Further Reading

Andrew, Christopher M. and Oleg Gordievsky. *KGB: The Inside Story of its Foreign Operations from Lenin to Gorbachev*. New York: HarperCollins-Publishers, 1990.

Fried, Richard M. *Nightmare in Red: The McCarthy Era in Perspective*. Oxford: Oxford University Press, 1991.

Hiss, Alger. *In the Court of Public Opinion*. New York: HarperCollins, 1957.

Klehr, Harvey, John Earl Haynes, and Fridrikh Igorevich Firsov. *The Secret World of American Communism*. New Haven, CT: Yale University Press, 1996.

Ruddy, T. Michael. *The Alger Hiss Espionage Case*. Belmont, CA: Wadsworth, 2005.

Swan, Patrick, ed. *Alger Hiss, Whittaker Chambers, and the Schism in the American Soul*. Wilmington, DE: Intercollegiate Studies Institute, 2003.

Weinstein, Allen. *Perjury: The Hiss-Chambers Case*. New York: Random House, 1997.

White, G. Edward. *Alger Hiss's Looking-Glass Wars: The Covert Life of a Soviet Spy*. Oxford: Oxford University Press, 2005.

13

John F. Kennedy was elected U.S. president in 1960 only because of voter fraud committed by his connections in the mafia.

PRO	Christian Nuenlist
CON	John H. Barnhill

PRO

John F. Kennedy, America's 35th president, still fascinates today, more than 45 years after his death. Even though his presidency was, with little more than 1,000 days, one of the briefest in U.S. history, the young, mediagenic, and charismatic Kennedy conveyed the image of a strong, powerful president with excellent leadership skills. Kennedy recalled the idealism of America's formative years and inspired a whole generation to perform public service passionately. Yet, the memories of the Kennedy era began to mix with myths immediately after the fateful November 22, 1963. The drama of a young hero, slaughtered under tragic and mysterious circumstances, shaped the early hagiographic historiography written by Kennedy's comrades-in-arms (Schlesinger 1965; Sorensen 1965). From the 1970s, revisionism brought harsh criticism: Kennedy's style had dominated his presidency, there had not been any substance, some historians wrote. The image had always been more important than concrete results, others added. I. F. Stone characterized the Kennedy era disrespectfully as an "optical illusion" (Leuchtenburg 1983: 53).

In 1997, award-winning investigative journalist Seymour Hersh summarized the entire revisionist criticism on Kennedy in a controversial but well-researched book titled *The Dark Side of Camelot*. Sunshine boy Kennedy, deeply loved by the public still today, mutated into the "Ugly American" par excellence. Hersh's Kennedy authorized murder conspiracies against Cuban leader Fidel Castro. Hersh's Kennedy, with the support of the secret service and loyal advisers, cheated on his wife in the White House swimming pool and elsewhere. Kennedy's frivolous private life, Hersh concluded, led to risky maneuvers even in the domestic and foreign policy of his government.

Yet, the most interesting part of the Kennedy saga actually is its very beginning. The presidential election of 1960 was one of the most close and tight races in U.S. history. In a truly cliffhanger election night, Kennedy beat his republican

287

rival Richard Nixon by merely 118,574 of more than 68 million votes. Kennedy's razor-thin victory was only known in the early morning hours of a dramatic election night, after fewer than 9,400 votes had brought the state of Illinois into the Kennedy camp. In the Electoral College, Kennedy won 303 votes against Nixon's 219 votes. The 1960 presidential election has been called "the most exciting democratic election of the twentieth century" (Aitken 1993: 274) and "the most famous presidential campaign of modern times" (Wicker 1991: 25).

Kennedy's narrow victory has triggered one of the biggest controversies in world history. Still today, critical journalists and historians doubt Kennedy's victory, accusing him of voter fraud. The target of these accusations is less the 35th president but rather a few illustrious actors who allegedly helped the young candidate to win in an illegal way, including his father Joseph P. Kennedy, Chicago's legendary Mayor Richard Daley, Chicago mafia boss Sam Giancana, and even the famous entertainer Frank Sinatra.

This section is structured into three parts. First, I focus on the important influence of Joseph P. Kennedy on the political career of his second-born son John Fitzgerald. Second, I analyze the controversial Kennedy victory at the West Virginia primary in May 1960—arguably the most important step in Kennedy's race for the democratic ticket. And finally, I delve into the presidential election in Chicago in November 1960. Almost 50 years after that crucial election, I argue, we have come to accept the distressing fact that the very close election victory of John F. Kennedy was only possible thanks to mafia votes.

Joseph P. and the Making of a President

Kennedy's father, Joseph P. Kennedy, was born in the United States in 1888 as a son of an Irish immigrant son. After World War I, he accumulated a fortune playing the stock market. He also became a player in the movie business and in illegal alcohol trade. Officially, he had a license to import medical alcohol (Whalen 1964; Schwarz 2003). Yet, mafia boss Frank Costello confided in 1973 that he and Joe Kennedy had been "partners" in the bootleg liquor business during prohibition (Hersh 1997: 44–57).

In 1931, the elder Kennedy became an important sponsor and fundraiser of Franklin D. Roosevelt. After three disastrous years serving as FDR's ambassador to Britain, he withdrew from public service after a bitter argument with the president. From that moment on, he knew that he would never become U.S. president. Instead, he began to force the political careers of his nine children. His oldest son Joe Jr. was slated for fulfilling his dream of a Kennedy presidency. When the navy pilot died in August 1944 during a secret bombing mission, the second-oldest son John F. had to step into Joe Jr.'s shoes. On Christmas of 1944, the ambitious father called John to take over the place of his older brother and to enter politics (Dallek 2003: 118).

In 1946, the 11th congressional district of Massachusetts in East Boston, where many Catholics lived, was targeted as the perfect launching pad of Kennedy's political career. Joseph Kennedy directed the campaign of his son; he developed the strategy and made almost every financial and political decision. The former ambassador called his influential friends, politicians, and journalists, for hundreds of hours. To veil his role, Joseph Kennedy usually paid for newspaper ads and other services in cash, mostly through an intermediary—all in all approximately $300,000. On June 18, 1946, the sickish and inexperienced 29-year-old John F. Kennedy won the democratic primaries and appeared on the front page of the *New York Times*. On November 5, 1946, he was elected congressman in the heavily democratic district (Bly 1996: 75–6).

In 1952 John F. Kennedy started the race for a Senate seat of Massachusetts—against the incumbent Senator Henry Cabot Lodge, Jr. Again, Kennedy's father helped discreetly, but efficiently. A few days before Election Day, Joseph P. learned that the *Boston Post*, a daily newspaper with a circulation of over 300,000 copies, badly needed money. Its publisher, John Fox, had favored the republicans and Cabot Lodge, but two days before Election Day, after a private meeting with Joseph Kennedy, Fox published a front-side endorsement for John F. Kennedy. In 1958 Fox admitted that the elder Kennedy had helped him out of financial difficulties with a loan of $500,000, which Fox later repaid. On November 4, 1952, Kennedy won the election (Bly 1996: 87).

During Kennedy's time as senator in Washington, father Joe continued to work on the jigsaw of a Kennedy presidency. As a former media mogul he recognized that his son was to campaign as a popular star and not simply as a normal politician. He paid $75,000 to his longtime friend and *Time* publisher Henry Luce for a front-cover story about the "Democrat's Man Out Front" on December 2, 1957. Kennedy for the first time smiled from a magazine cover, even though his political achievements in the U.S. Senate actually were rather mingy. Kennedy was by many considered to be "more the product of a public relations campaign funded by his family's fortune than the result of political accomplishments" (Dallek 2003: 230).

West Virginia—Nomination for Sale

On January 2, 1960, John F. Kennedy finally launched his presidential campaign. Again, it was a family business. His younger brother Bobby worked as campaign manager; father Joe helped as usual in the background as hidden paymaster. On the way to the democratic nomination, the primary in West Virginia on May 10 was the biggest hurdle. Kennedy's most dangerous contender in the inner-democratic race at that time was Hubert Humphrey, a senator from Minnesota.

One particularity of West Virginia county-level politics was (and still is today) the custom of "slating." A slate is a printed list of candidates. If a candidate

---

The Making of the President 1960

The 1960 presidential election, which saw John F. Kennedy becoming the 35th U.S. president, was the title of the first of five books written by journalist Theodore H. White on American politics. *The Making of the President 1960* was a brilliantly written story of an election, which, even though everybody knew the final outcome, left the reader in suspense until the vote was counted.

Theodore White (1915–1986) was born in Boston, the son of a lawyer, and studied at Harvard University, being a classmate of Joseph P. Kennedy, Jr. (older brother of John F. Kennedy). After becoming a journalist in China, and a friend of Henry Luce, he wrote a number of books on China.

In his book *The Making of the President 1960*, he gives detailed coverage of the entire campaign, but also a scientific analysis of aspects of it. Awarded the Pulitzer Prize for general nonfiction in 1962, White went on to write *The Making of the President 1964* (1965), *The Making of the President 1968* (1969), *The Making of the President 1972* (1973), and *Breach of Faith: The Fall of Richard Nixon* (1975).

---

became associated with the "right" slate, his chances of winning were improved. Sheriffs or political committeemen in the 55 counties, who put together the lists or slates, tightly held political control. Political tradition called for statewide candidates to pay some or all of the county's election expenses in return for being placed at the top of a slate. Paying a few dollars per vote on Election Day was widespread in the state (Fleming 1992: 13–4).

Humphrey's defeat in West Virginia was "unexpected, total, and humiliating" (Griffith 1965: 17). A Humphrey victory had been widely predicted (Sorensen 1965: 146). Alan L. Otten, the *Wall Street Journal* correspondent covering the campaign in West Virginia, thought that Humphrey would capitalize on the pronounced anti-Catholicism in the state and win the primary handily. "Every miner I talked to was going to vote for Humphrey," he recalled later (quoted from Hersh 1997: 92). When Kennedy's victory was announced, Otten was convinced that the votes had been bought by the Kennedys and started an investigation. After five weeks, the *Wall Street Journal*'s investigative team learned that the Kennedys had turned what had historically been random election fraud into a statewide pattern of corruption and had stolen the election from Humphrey. The reporters concluded that huge sums of Kennedy money had been funneled into West Virginia, much of it from Chicago. In the end, however, there was no smoking gun, none of the newspaper's sources saw a representative of the Kennedy campaign give money to a West Virginian, and thus the editors in Washington ruled that the article could not be published.

Seymour Hersh, however, learned from interviews with county and state officials that the Kennedy family spent $2 to 4 million in bribes and other payoffs, with some sheriffs in key counties collecting more than $50,000 apiece in

cash in return for placing Kennedy's name at the top of their election slate. According to Chicago coal buyer James McCahey, the political payoffs in West Virginia had begun in October 1959, when Edward "Teddy" Kennedy traveled across the state distributing cash to the democratic committeeman in each county. With the payoffs amounting to $5,000 per committeeman, the Kennedy camp spent $275,000. In a later stage of Kennedy's West Virginia campaign, when new polls showed a dramatic drop in support for Kennedy among West Virginians, McCahey learned at a strategy meeting of the Kennedy campaign of Teddy Kennedy's 1959 mission. He lectured the outsiders that the Kennedys had to target the local sheriffs, not the committeeman. He recommended to organize teachers and other grassroots workers and later acknowledged passing out some cash to local political leaders, paying as much as $2,000 for storefront rentals and for hiring cars to bring voters to the polls on primary day. According to Victor Gabriel, who ran the Kennedy campaign in Harrison County, Bobby Kennedy showed him a little black book that listed $40,000 as being given to Sid Christie, an attorney who was the top democrat of McDowell County in the heart of the state's coal belt in the South. Two former state officials acknowledged during interviews with Seymour Hersh that some sheriffs had been paid as much as $50,000 (Hersh 1997: 96ff.).

Clearly, the huge amount of money passed by the Kennedy camp to sheriffs and local officials helped Kennedy to win the crucial state. But what was the role of the mafia, and how was organized crime involved in supplying the cash that enabled Kennedy to win the West Virginia primary? There were links between entertainer Frank Sinatra, mobster Sam Giancana, Joseph Kennedy, and John F. Kennedy. Sinatra and his friends in the "Rat Pack," including Dean Martin and Sammy Davis, Jr., raised money for the Kennedy campaign. Memoirs and Federal Bureau of Investigation (FBI) wiretaps confirm earlier suspicions and accusations by the press that under-the-table mafia money disbursed through Sinatra was used to make payoffs to key election officials (Kelley 1987: 295; Summers and Swan 2006: 271).
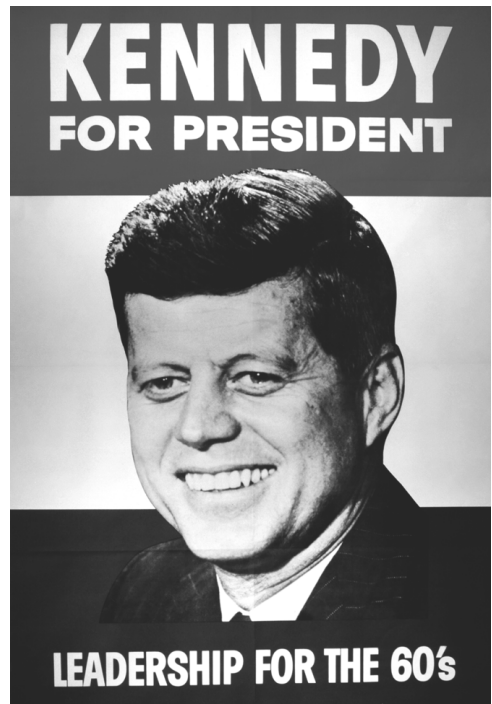
A campaign poster for John F. Kennedy's successful presidential bid in 1960. (John F. Kennedy Library)

The exact amount of money, however, is difficult to trace. Paul "Skinny" D'Amato, a New Jersey nightclub owner, said that he was approached by Joe Kennedy during the primary campaign and asked to raise money for West Virginia (Van Meter 2003). In return, the elder Kennedy promised that President John F. Kennedy would reverse a 1956 federal deportation order for Joey Adonis, a New Jersey gang leader. D'Amato allegedly raised $50,000—an amount that was too small to decide the election if compared with the sum that the Kennedy campaign spent in West Virginia. Thus, it seems fair to conclude that while Mafia money helped John F. Kennedy win the West Virginia primary against Hubert Humphrey, the crucial factor was simply the enormous amount of money the Kennedys pumped into the campaign. Clearly, they bought the election. D'Amato did not get anything in return for his $50,000. When Attorney General Robert Kennedy was informed in early 1961 by the FBI that D'Amato had bragged about his role in moving cash from Las Vegas to help John F. Kennedy win the election, the mobster suddenly faced federal indictment on income tax charges from his failure to file a corporate tax return for his nightclub (Hersh 1997: 100–01).

Chicago—The Stolen Election

Once John F. Kennedy had eliminated Humphrey in the West Virginia primary and sealed the democratic ticket in Los Angeles in June, the Kennedy camp prepared for the fight against the republicans. On Election Day, on November 8, 1960, Kennedy defeated Richard Nixon by 118,574 votes and 303 electoral votes—34 more than he needed to win. It was a very narrow victory. Charges of vote fraud appeared in several states, and the republicans sent officials on troubleshooting missions to seven states to detect vote fraud. The election in Illinois was the most controversial. The state of Illinois was essential to Kennedy's victory: Without the state's 27 electoral votes, Kennedy would have had only 7 votes more than necessary to win the Electoral College, with 26 unpledged democratic electors in Mississippi, Georgia, and Alabama. If Kennedy had lost Illinois, these delegates would have had the power to throw the election into the House of Representatives.

There were widespread rumors about vote fraud in Chicago—Kennedy won the city by 456,312 votes, with more than 89 percent of the eligible voters having voted (or recorded as having voted). The huge plurality in Chicago enabled Kennedy to offset a wave of downstate republican votes. Chicago's legendary Mayor Richard Daley told Kennedy on election night: "Mr. President, with a little bit of luck and the help of close friends, you're going to carry Illinois" (quoted from Bradlee 1984: 33).

Who exactly were these "helping friends"? Joe Kennedy knew that organized crime dominated the major unions in Chicago in the 1950s and 1960s. Chicago's Mafia offered a huge manpower base that could be mobilized on demand and the Kennedy patriarch was ready to use that help. And old friend, William J.

Tuohy, chief judge of the Circuit Court of Cook County (Chicago), was asked to arrange a secret meeting with Sam Giancana. Tuohy approached Giancana through Robert J. McDonnell, then one of the mob's leading attorneys and a former protégé of Tuohy. According to McDonnell, he was involved in setting up this delicate meeting in the winter of 1959–1960. In the 1990s, McDonnell declared: "I'm convinced in my heart of hearts that Giancana carried the day for John F. Kennedy" (Hersh 1997: 135–36). The Chicago mafia put drivers in every precinct to help out the precinct captains to get the voters out. And the mob had the unions voting for Kennedy.

A second source for the Kennedy-Giancana meeting appeared in 1992 in a CBS series produced by Tina Sinatra. Her father, the famous entertainer Frank Sinatra, provided the information for the scenes and explicitly approved the script. For the show of his daughter, Sinatra broke his long-held secret and publicly admitted for the first time that he had played a role in brokering Giancana's support for John F. Kennedy in 1960. According to Tina Sinatra, Joe Kennedy invited her father to a meeting in late 1959 at Hyannis Port. Over lunch, the elder Kennedy asked Sinatra to help him in West Virginia and Illinois "with your friends." Frank Sinatra agreed and met Sam Giancana on a golf course, away from FBI surveillance, and asked him to ensure the wins in Illinois and West Virginia by getting the numbers out and getting the unions to vote (Hersh 1997: 137–38).

J. Edgar Hoover's FBI picked up hints of the preelection bargaining between Kennedy and the Chicago mafia, but could not make them public because FBI agents had begun installing bugging devices in the late 1950s without specific court authorization to do so. In November 1960, the FBI heard Giancana talk about an election deal with Joe Kennedy. In his 1989 memoir *Man against the Mob*, FBI agent William F. Roemer, Jr., revealed that Giancana had been overheard on a still-unreleased FBI wiretap discussing an election deal: mafia support in return for a commitment from the Kennedy administration to back off from the FBI investigation of Giancana (Hersh 1997: 140).

Giancana also bragged about his influence to Judith Campbell Exner, the Los Angeles woman who was sexually involved with John F. Kennedy while also meeting with Giancana. In *My Story*, Judith Exner's 1977 memoir, the mob leader is quoted saying to her: "Listen, honey, if it wasn't for me your boyfriend wouldn't even be in the White House" (Exner 1977: 194).

Another source for the all-out support of the Chicago syndicate for Kennedy is G. Robert Blakey, a former special prosecutor for the Justice Department. Blakey, who had access to still classified FBI wiretaps, said: "The FBI bugs in Chicago demonstrate beyond doubt that enough votes were stolen in Chicago to give Kennedy a sufficient margin that he carried the state of Illinois." Blakey added that the surveillance in Chicago also established that mafia money was put in the 1960 national election. The money traveled from singer Frank Sinatra, who was a close friend of Giancana, to Joe Kennedy, Blakey said. In return,

Giancana expected that the Kennedys would reduce FBI pressure on their activities (Hersh 1997: 140).

After the election, Everett Dirksen, the ranking republican senator from Illinois, requested a full-scale FBI investigation of the election. He claimed to have evidence that the election had been stolen. The FBI told him that they had already received considerable information on vote fraud and that they were sending this information to the Department of Justice. Of course, the department, headed by presidential brother Robert F. Kennedy, did not act on the information and did not conduct any further investigation (Hersh 1997: 154).

FBI Director J. Edgar Hoover, of course, also knew that Giancana had provided help for the Kennedy campaign in Illinois and provided money for the crucial West Virginia primary. Hoover told Kennedy about it in November 1961, when he learned from FBI bugs about it. Kennedy told his father, and within a few days, Joseph Kennedy had a massive stroke from which he never fully recovered (Gentry 1991: 472).

Conclusion

Fifty years after John F. Kennedy's razor-thin victory over Richard Nixon, it seems clear that money, muscle, and influence of organized crime helped Kennedy win the 1960 presidential election—the closest election in the 20th century and the election with the highest voter participation in the century. Kennedy's margin of victory was just 118,574 votes. The young Democratic front-runner for the presidency was not a typical presidential candidate. His career in the U.S. Senate had been unimpressive. He sponsored no major legislation, had a poor attendance record, and seemed bored by senatorial duties. Instead, he rose to national prominence in 1956 when his book *Profiles of Courage* won the Pulitzer prize and when he almost won the vice presidential nomination at the Democrat National Convention. His father, Joseph P. Kennedy, one of the wealthiest men in America, largely financed his presidential campaign.

The most intriguing aspect of the buying-the-election charge against Kennedy surfaced in 1979, growing out of the House of Representative hearings concerning the assassination of President Kennedy. Drawing from FBI files, some fascinating information was uncovered, particularly from wiretaps of Sam Giancana, the leading Midwestern mafia overlord. According to this information, the mafia was brought into the Kennedy presidential campaign and critically intervened at two crucial stages—at the primary in West Virginia in May 1960 and at Election Day in Chicago in November 1960.

Mafia money bought votes in the May 10 primary in West Virginia, although it is impossible to determine in retrospect whether the mob involvement was "the deciding factor in the contest," as John Davis has argued (Davis 1984: 286). All evidence points to mafia donations to Kennedy's West Virginia

campaign. Skinny D'Amato, for example, spent $50,000 and was promised by Joseph P. Kennedy a favor in return—a favor that President John F. Kennedy and his brother Robert F. Kennedy as attorney general did not implement during the Kennedy presidency. Compared with the huge amount of money that the Kennedys pumped into the primary campaign—$2 to 4 million—the mafia money probably did not make the crucial difference for Kennedy to win, even though the West Virginia primary was clearly bought.

Joseph P. Kennedy did more than invest his time and money into the presidential campaign of his son. He also made a risky bargain with rather dubious mobsters, including, most prominently, with Sam Giancana and the powerful organized crime syndicate in Chicago. To ensure Kennedy's victory in West Virginia and Illinois, the elder Kennedy arranged a personal meeting with Giancana. The deal they made included an assurance that Giancana's men would get out the Kennedy vote in the mob-controlled unions in Chicago and elsewhere. Thus, John and Bobby Kennedy took office in January 1961 knowing that the mafia had helped win the 1960 election. Giancana felt optimistic that by backing a presidential winner, the heat would be off the Chicago syndicate. Charges of voter fraud were also directed at Chicago Mayor Richard Daley who had personally promised to deliver Illinois to Kennedy. Richard Dallek, one of the most respected U.S. presidential historians, wrote in 2003: "Daley's machine probably stole Illinois from Nixon" (Dallek 2003: 295).

Richard Nixon knew what had happened in Chicago and was urged by republican leaders to demand a recount. In his memoir *RN*, Nixon admitted that his decision not to contest the legitimacy of the election—one of his finest hours according to many Americans—was based on self-interest: "And what if I demanded a recount and it turned out that despite the vote fraud Kennedy had still won? Charges of 'sore loser' would follow me through history and remove any possibility of a further political career" (Nixon 1978: 224).

The result of the 1960 election was so close that almost any adverse factor can be blamed for Nixon's defeat. Nixon himself was convinced that "to ascribe defeat or victory to a single factor in such a close contest is at best guesswork and oversimplification" (Nixon 1978: 221). Of course, it was not just the money and the mafia that lead to the narrow Kennedy victory. There were other factors as well. Cortisone injections gave Kennedy a healthy appearance and movie-star good looks. Although he was only four years younger than his republican rival, Kennedy capitalized on the contrast to President Dwight D. Eisenhower, who was 27 years older, and evoked the promise of a young, activist leadership ("Let's get this country moving again!"). The image of an athlete, war hero, and a devoted family man that Kennedy and his handlers created was a winning formula. In reality, Kennedy was a very sick man with terrible back pain; *Profiles in Courage* had been ghostwritten; his World War II record was courageous, but spotty; and his philandering was almost insatiable (but publicly unknown at that

time). Yet, the image worked, in particular in contrast to Nixon and because the 1960 race was the first presidential election dominated by television. This new medium helped emphasize style over substance, and 70 million Americans watched the televised debates between Kennedy and Nixon.

In the end, however, none of the factors that lead to John F. Kennedy's narrow victory at the U.S. presidential elections in 1960 was as critical and dangerous as the involvement of the mafia in the campaign. Clearly, the mafia felt they had delivered what they had promised and expected favors by the new U.S. government in return. The Kennedy dynasty was aware of the deal with the mafia. When John F. Kennedy in the weeks before his inauguration complained about the difficulties in finding top people for the key posts in his administration, Joe Kennedy replied: "Jack, if you don't want the job, you don't have to take it. They're still counting votes up in Cook County," referring to the mob-dominated West Side of Chicago that provided the crucial votes for winning the state of Illinois (Bly 1996: 75).

Once the two Kennedy brothers were in charge, however, they began a concerted effort to wage a war on organized crime. Hundreds of mafia figures were prosecuted and convicted. The mafia was angry and felt double-crossed by the Kennedys. But the mobsters might have had their revenge. Although the official version of the Kennedy assassination points to Lee Oswald as lone murderer, in recent years new evidence has come up that hints at mafia involvement (Bugliosi 2007; Kroth 2003; Talbot 2007).

References and Further Reading

Aitken, Jonathan. *Nixon: A Life*. London: Weidenfeld, 1993.

Bly, Nellie. *The Kennedy Men: Three Generations of Sex, Scandal*. New York: Kensington, 1996.

Bradlee, Benjamin C. *Conversations with Kennedy*. New York: Norton, 1984.

Bugliosi, Vincent. *Reclaiming History*. New York: Norton, 2007.

Dallek, Robert. *An Unfinished Life: John F. Kennedy 1917–1963*. Boston: Little, 2003.

Davis, John H. *The Kennedys: Dynasty and Disaster*. New York: McGraw-Hill, 1984.

"Democrats: Man Out Front." *Time* (December 2, 1957): 1, www.time.com/time/magazine/article/0,9171,825326-1,00.html (accessed June 24, 2010).

Exner, Judith, *Judith Exner: My Story*. New York: Grove, 1977.

Fleming, Dan B., Jr. *Kennedy vs. Humphrey, West Virginia, 1960: The Pivotal Battle for the Democratic Presidential Nomination*. Jefferson, N.C.: McFarland, 1992.

Gentry, Curt *J. Edgar Hoover: The Man and the Secrets*. New York: Norton, 1991.

Giancana, Sam, and Chuck Giancana. *Double Cross: The Explosive Inside Story of the Mobster Who Controlled America*. New York: Warner, 1992.

Griffith, Winthrop. *Humphrey: A Candid Biography*. New York: Morrow, 1965.

Hersh, Seymour M. *The Dark Side of Camelot*. Boston: Little, Brown, 1997.

Kelley, Kitty. *His Way: The Unauthorized Biography of Frank Sinatra*. New York: Bantam Books, 1987.

Kroth, Jerome A. *Conspiracy in Camelot. The Complete History of the Assassination of John F. Kennedy*. New York: Algora, 2003.

Koskoff, David E. *Joseph P. Kennedy: A Life and Times*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

Leuchtenburg, William E. "John F. Kennedy Twenty Years Later." *American Heritage* 51, 3 (December 1983): 51–59.

Nixon, Richard M. *RN: The Memoirs of Richard Nixon*. London: Sidgwick, 1978.

O'Donnell, Kenneth P., and David E. Powers. "*Johnny, We Hardly Knew Ye": Memories of John Fitzgerald Kennedy*. Boston: Little, Brown, 1970.

Royko, Mike. *The Boss: Richard J. Daley of Chicago*. New York: Dutton, 1971.

Schlesinger, Arthur M., Jr. *A Thousand Days: John F. Kennedy in the White House*. Boston: Houghton Mifflin, 1965.

Schlesinger, Arthur M., Jr. *Robert F. Kennedy and His Times*. Boston: Houghton Mifflin, 1978.

Schwarz, Ted. *Joseph P. Kennedy: The Mogul, the Mob, the Statesman, and the Making of an American Myth*. Hoboken, NJ: Wiley, 2003.

Sorensen, Theodore. *Kennedy*. New York: Harper and Row, 1965.

Summers, Anthony, and Robbyn Swan. *Sinatra: The Life*. New York: Vintage, 2006.

Talbot, David. *Brothers: The Hidden History of the Kennedy Years*. New York: Free Press, 2007.

Van Meter, Jonathan. *The Last Good Time: Skinny D'Amato, the Notorious 500 Club, and the Rise and Fall of Atlantic City*. New York: Crown, 2003.

Whalen, Richard. *Founding Father: The Story of Joseph P. Kennedy*. New York: New American Library, 1964.

White, Theodore H. *The Making of the President, 1960*. New York: Atheneum, 1961.

Wicker, Tom. "One of Us." In *Richard M. Nixon: Politician, President, Administrator* (pp. 23–34). Edited by Leon Friedman and William F. Levanstrosser. Westport, CT: Greenwood, 1991.

CON

Although Seymour Hersh's 1997 *The Dark Side of Camelot* is the fullest development and grimmest, it is not the sole theory that the Kennedy ties to the mob were pivotal in the election of 1960. Since the first appearance of the Camelot myth after the Kennedy assassination, there has been at least an undercurrent of myth breaking. One way to destroy the illusion of Camelot is to keep alive the theory that as a favor to Joe Kennedy, founder of the Kennedy fortune and political success, the mafia capo Sam Giancana used his control of the Cook County unions, and specifically the teamsters' pension fund, to alter the results of the presidential election of 1960. Ballot box manipulation and goon intimidation were among the methods that Giancana employed in expectation that a grateful President Kennedy would ease the federal pressure on the mob, and Giancana in particular. All his work succeeded in giving John Kennedy a 10,000-vote margin in the county, sufficient to give Kennedy a 9,000-vote edge in Illinois, gaining Illinois's electoral votes for Kennedy and taking the election from Richard Nixon.

The Giancana-Kennedy relationship, subscribers to this theory contend, took a turn after the election when the two men had a common mistress, Judith Campbell Exner, who served as liaison between them and as courier for large sums of money and sensitive information, including plots to assassinate Fidel Castro of Cuba.

The argument is provocative, but it does not fit the facts. Not only is there no evidence other than hearsay, but the principals are conveniently dead and unable to defend themselves. More important, if there was a fix, it didn't take. Vote totals in the precincts where the alleged manipulation occurred were consistent with those in precincts where the customary Chicago standard applied. Rather than the mob handing the election to Kennedy, a stronger case can be made that in fact Kennedy won on his merits.

The earliest proponents of the theory, Brashler in 1977 and Roemer in 1989, half-stepped it, giving the major role to Richard Daley and his democratic machine and limiting mafia activity to a few wards. Sam Giancana's half-brother and nephew, Chuck and Sam Giancana in 1992, Hersh in 1997, and Russo in 2001 took a more extreme position, contending that the mob effort was huge and either brought out the union vote in the states where mob influence was significant or brought out a tainted union vote throughout the country or in the narrow view influenced Chicago to Kennedy's advantage. Hersh and the Giancanas agree that Joe Kennedy, the former bootlegger, used his mob ties to make a deal with Giancana. Chuck and Sam Giancana's *Double Cross* (1992) says that Joe

Kennedy got in touch with Sam Giancana and after several meetings the two made a deal—elect John Kennedy and the heat comes off.

Sam and Chuck Giancana say that the outfit engaged in massive vote fraud and had hoodlums intimidating voters to the point of breaking limbs of those who objected to voting for Kennedy. The Kennedys double-crossed the mob by increasing pressure on the Chicago mafia rather than backing off. The Giancanas go further in arguing that the new administration double-crossed Giancana by intensifying its crusade against the mob, and the mob retaliated by killing the president and his brother.

Although he dismisses the idea that the mob would have done away with Kennedy and indicates they were as surprised as anyone, Gus Russo reaffirms the Giancanas and Hersh. He says the mob used extreme measures, muscle, and money in Chicago. Chicago's Murray "The Camel" Humphreys worked the mob-influenced unions nationally, and the involved national unions included some nonteamsters. He says the four states the mob delivered were Illinois, Michigan, Missouri, and Nevada. Russo finds the Chicago mob, the "Outfit," to be a serious business whose members put in long hours but did not take their work home with them.

The Flaws

Witnesses

So how good is the evidence? Who are the witnesses? How reliable are they? Seymour Hersh says that Kennedy did not deal with Giancana directly, weakening the contention that Kennedy was using mob ties from the prohibition era. Rather, Joe asked Judge William Tuohy to help get a message to Giancana. Supposedly the judge used Robert J. "Bob" McDonnell to arrange the meeting between Kennedy and Giancana in the judge's chambers. The source for this bit of information is McDonnell, and it does not hold up because the judge didn't need McDonnell, who didn't know Giancana. If Tuohy wanted to contact Giancana, he had contacts in the First Ward, John D'Arco and Pat Marcy, who were close to Giancana and could easily arrange a meeting. And why would anyone want McDonnell to witness the meeting? McDonnell claims he watched the introduction of Kennedy and Giancana, leaving with Tuohy immediately thereafter.

McDonnell has credibility problems. He was disbarred and drank and gambled compulsively, borrowed from a mob loan shark, and ended up running errands for the mob, including moving a couple of corpses. He was unstable and unreliable and not exactly the sort of person that the close-mouthed principles would have requested to not only arrange but witness a meeting with the potential, if revealed, of destroying the Kennedy political future. McDonnell claims the mob turned out ward votes and union votes at some unclear level.

The wife of Murray Humphreys, in 1960 the mob's go-to guy on political corruption, is another source. She says she saw her husband coordinating the national effort and helped him. Mafia wives are not privy to the business. Union and mob leaders wouldn't do business in front of any woman, wife or girlfriend, and certainly wouldn't have a woman helping with their work.

Then there is the coauthor of *Double Cross*. Chuck Giancana was a low-level mobster who would not be privy to much of what the close-mouthed upper echelon dealt with. The mob operated on a need-to-know basis anyway. And the Giancanas' book blames Sam for every crime from his early teens, even crimes that evidence pins on others.

Most of all, there is Judith Exner. Taken as true, based on Federal Bureau of Investigation (FBI) and White House records, is the fact that she had an affair with John Kennedy and a relationship with Sam Giancana. The documents support the story that Exner had a brief romance with Sinatra in 1959. Frank Sinatra introduced her to John Kennedy at a "Rat Pack" gathering in Las Vegas during the presidential campaign of 1960. Soon the two began their affair.

A few months later Sinatra introduced her to "Sam Flood," whom she later learned was Sam Giancana, head of the Chicago mob.

Exner and John Kennedy had a three-year affair that included trysts in the White House and in the Kennedy's Georgetown house. Exner was seeing Giancana while seeing John Kennedy, but she claims the relationship was that of friends only until after she split with John Kennedy (in another version she and Giancana were in a relationship within a month).

J. Edgar Hoover gave John Kennedy the FBI files on the Exner-Giancana relationship in March 1962. The relationship came out in the Church Committee's 1975 report, which linked Exner to not only Giancana but also to Johnny Roselli, another mobster. Revealed, Exner went public in admitting the affair. She denied, however, that she knew anything about mob activities.

Beyond that, there are problems with the reliability of Exner's recall of events. Exner denied three times that she had any idea of a relationship between John Kennedy and Giancana, first in the Church Committee and next in a news conference in 1975 and again in her book of 1977. At that time even the affair brought skeptical responses, but over time as knowledge of John Kennedy's womanizing emerged, Exner's story became more credible. Exner claimed that Giancana wanted to marry her, and she refused to talk about Giancana and Roselli to the FBI because by then she'd already moved on to Eddie Fisher and an unnamed pitcher for the Angels.

She faded from the scene until 1988 when Kitty Kelly published a *People Magazine* story in which Exner claimed she had earlier lied to the Church Committee because she was afraid after Giancana was killed in 1975, and Roselli in 1976, with neither murder solved. Now she said she had been a regular go-between, meeting with the two gangsters often at Kennedy's request, carrying

documents and envelopes over a two-year period from the spring of 1960. She claimed she set up a meeting between Giancana and Kennedy. She claimed the only reason she had a relationship with Giancana was as a favor to Kennedy. She claimed that Bobby Kennedy knew she was a courier.

Again in 1996 she reappeared, this time for an interview with Liz Smith in *Vanity Fair*. The story changed again, and at this point she becomes a direct witness for the Chicago fix story because she said that in April 1960 Kennedy had told her he was giving Giancana money to help with the primaries. She said he showed her the money in an envelope she was to deliver. And he told her what was in the other envelope, the Castro plot documents. She said that Bobby was involved and that she carried bribes from California contractors to the president and that she aborted Kennedy's child. Smith dismissed the Exner stories as desperate publicity seeking by an alcoholic cancer patient.

Less skeptical than Smith, the next year, Hersh used Exner in *The Dark Side of Camelot*. Hersh said there was a witness, Martin Underwood, who initially said he followed her on the train in April 1960 to make sure she delivered the money. Underwood recanted.

The Exner story has a few problems. Why would John Kennedy meet with mobsters when in all likelihood he knew that the FBI was bugging them? That's what the underlings were for. And his brother Bobby was at war with the mob. Was John Kennedy the sort to trust a casual girlfriend with large sums of money and highly sensitive documents—and to collect payoffs from General Dynamic, another allegation in the Exner story as mentioned by Hersh? Exner said that Kennedy did not trust the Central Intelligence Agency (CIA) and felt that a woman could get through because no one would think that she could do such a job.

For skeptics, Exner is unreliable, has a history of depression as well as addictions to amphetamines and alcohol, and has suffered harassment at the hands of the FBI. One interpretation is that she wanted to improve her image after the initial story led to her being labeled a party girl or call girl. Being a courier between two extremely powerful men is a better story than being bedded by a bunch of questionable men. For self-protection, she imagined herself a spy out of a best-selling thriller. The new version also matched the current conspiracy stories that tied the mob to John Kennedy's death and to a CIA effort to assassinate Castro. But the motivation for her changing story is secondary. The point is that the story changed over time.

And there are the witnesses whose testimony isn't included by the mob-fix theorists. The initial reaction to *The Dark Side of Camelot* by former Kennedy insiders was negative. One reviewer described the book as a series of exposés in a relentless effort to demythologize the president from Camelot. Theodore Sorensen labeled the work a compilation of all sorts of wild stories, and Arthur Schlesinger, Jr., regarded it as suitable for the gullible author or reader, and noted that his White House didn't have a bimbo parade.

The Mob

Chicago investigative reporter Len O'Connor claimed as early as 1975 that the mob in 1960, rather than being in bed with one Kennedy or another, was in fact hostile to Kennedy. By then organized crime had suffered federal heat for three years, dating from the exposure of the Appalachian meeting of 1957 and the coincident hearings by the Senate McClellan Committee, including John and Robert Kennedy. The investigation included the business of Sam Giancana in Chicago. Robert Kennedy ridiculed Giancana at the hearing by comparing his laugh to a little girl's: "I thought only little girls giggled."

O'Connor did not rely on hearsay. Rather he compared 1960 presidential results with 1955 mayoral results in some Chicago wards and claims that the vote in the mob-controlled wards, rather than aiding Kennedy, was in fact unusually heavy against Kennedy. According to Len O'Connor (1975), the Daley Machine demonstrated its clout with only two wards giving a lower vote for Kennedy than for Daley in 1955. Those two were the 1st and 28th wards, both controlled by the mafia. Daley's people read this as a message from the mob that it disapproved of Bobby's treatment of them during the McClellan hearings. O'Connor also says that Murray Humphreys convinced his friend, the 45th ward alderman, to openly oppose Kennedy's effort. So the mob and the 45th worked against Kennedy, not for him. O'Connor agrees that mob-influenced unions were anti-Kennedy because of McClellan.

Chicago mafia boss Sam Giancana appears before a Senate committee on racketeering, June 9, 1959. (Bettmann/Corbis)

Brashler argues in his biography of Sam Giancana that Sinatra asked Giancana to help Kennedy, but that the mob's help didn't change the outcome, merely padded the victory that the Daley Machine locked up, being gung-ho for an Irish Catholic. The mob wards did basically the same as the Daley wards did.

And where is all that violence claimed by the Giancanas? Problems with the stories include the lack of any mention in any of Chicago's four daily newspapers of any violence during the election.

Also, the mob had control of only 5 of the 50 Chicago wards, so it didn't have the means to deliver the vote. The combined population of the 1st, 24th, 25th, 28th, and 29th wards in

1960 was 306,169. Even unusually large Kennedy votes in those five would not be enough to swing even the city much less the state. Even in those five wards, with only 300 members (and many of advanced age) the mob was too small to do the muscle work in the 279 precincts, each of them requiring, theoretically at least, five hoodlums for intimidation. Also, the control of the national teamsters doesn't wash. The mob families were territorial, so the best the Outfit could deliver was the Chicago locals. Jimmy Hoffa despised the Kennedys and backed Nixon, and odds are high that the Chicago teamsters would not have backed him.

The Evidence for a Fix?

When John Binder (2007) analyzed the ward totals for the supposed Giancana-controlled wards, he found that those areas voted no differently from other wards in the city. Democratic votes in 1960 were about the same as they were in 1956.

Joe Kennedy had money and spent it as lavishly as he spent his media ties with the single purpose of getting John elected. He began the process of getting press coverage in the mid-1950s, including a 1957 *Time* cover despite John's lack of accomplishment. Joe said the cover cost him $75,000. He also had William Bradford Huie handing out cash to politicians who could help John. Giancana and Sinatra canceled debts and handed out cash to West Virginians in the primary. And Massachusetts politicians descended on West Virginia with wads of money for favors. Joe also collected money through the Catholic Church of Boston, writing checks for more than an average Sunday collection, then taking the cash and using it without traceability. Cardinal Cushing claimed that he and Joe sat together and decided which preacher got how much money.

Joe Kennedy was a manipulator and he contacted many people, maybe including mob bosses, to make sure his son got the presidency in 1960. Maybe Kennedy did meet with Giancana, but there is no evidence that the outcome of the race was altered by the mobster. He controlled two wards, already heavily democratic, in a city that the Daley Machine fully controlled. The votes in these two wards were low compared to those in the other nine automatic democratic wards, all of which maxed out their possible vote for Kennedy.

The numbers do not match the argument for a fix. The wards and suburbs that the Outfit controlled and in which they supposedly got out the labor vote didn't vote any differently than normal. If the Outfit was involved, most likely it double-crossed the Kennedys by not delivering the votes it said it would. The mob had the capability of altering the outcome in some wards, but it had no reason to do so. In fact, there is no evidence that the mob-controlled wards did go heavily for Kennedy. The mob did use its clout against the republican incumbent state's attorney for Cook County Benjamin Adamowski, and straight ticket voting did spill over to Kennedy, so there was a small increase in

democratic support in the mob-controlled areas. It was not enough to warrant a claim that the mob used its clout on his behalf.

The Republican Reaction

The race was the closest of the 20th century, tight enough that only 113,000 votes of 68 million separated the two candidates. Texas, Illinois, and many other states were so close that they could have gone either way. For instance, Kennedy had California until the absentee ballots were recorded; then Nixon took the state. Nixon won Hawaii's three votes after a challenge, but that's all.

Stories of fraud began appearing even before the vote. Richard J. Daley was long known for delivering the vote, no matter what it took. And, although Nixon didn't contest the results, some of his supporters did. But they did not focus specifically on Illinois. They challenged results in 11 states. They got recounts, grand juries, and FBI investigations. They had their claims heard and evaluated based on the evidence. Not much came of it. In New Jersey, for instance, the state republicans found out early that their recounts were not showing any significant discrepancies. Similarly, recounts and investigations just sort of faded away in the other states. In Illinois the challenge was fierce and focused on Cook County, with its 450,000-vote Kennedy margin. The recount showed that although there were undercounts for Nixon, in 40 percent of the precincts the Nixon vote was overcounted. The republican-dominated

---

Controversial Close Elections in American History

Some U.S. presidential elections have been very close, and the 1960 election result was only one of several occasions when the result was not known for some time after Election Day. The most delayed announcement of a final winner in a presidential election was in 2000 with the U.S. presidential election between George W. Bush and Vice President Al Gore. This saw the votes in parts of Florida recounted many times.

Another close presidential election was that in 1948 between President Harry S. Truman and republican aspirant Thomas E. Dewey. In order to ascertain the result ahead of the final vote counting, a telephone poll was conducted by the *Chicago Tribune*, which then went to press with its famous headline 'Dewey Defeats Truman.'

And in 1888, the incumbent president, Grover Cleveland, lost the election to Benjamin Harrison, although Cleveland managed to win more votes than Harrison. However, Harrison managed to gain a majority in the Electoral College, a result that came from a surprise victory in New York, Cleveland's home state. Had the New York vote gone the other way, Cleveland would have won. Indeed four years later, Cleveland did win the next presidential election, returning to the White House.

State Board of Elections rejected the republican case, stating that the party had not given even one affidavit on its behalf.

This doesn't prove that fraud didn't exist. It just indicates that the republicans couldn't prove it. More convincing is the work of Edmund Kallina (1988), who reinforced a 1961 University of Chicago study that indicated that whatever fraud there might have been, it was not enough to alter the outcome.

Only three people went to jail for election-related crimes, and 677 others were acquitted, admittedly by a Daley henchman. For a presidential election in a major American city, those numbers are closer to business as usual than to a major fix of the election.

Numbers Support a Kennedy Win

At the time, the republicans blamed Daley and his machine for fixing the election. Daley did want to get out the vote to unseat a potential rival for his run for mayor, and he did get out the vote as well as unseated the rival. He also controlled Kennedy's appearances, keeping the candidate campaigning in the South to build anticipation until he was ready to bring him to Chicago. Then he brought him to a rally in Chicago Stadium after a torchlight parade, and talk that night was of a 500,000-vote promise in Chicago, 400,000 in Cook County by Daley to Kennedy. The drama of the evening electrified the faithful. Experts anticipated Cook County going by a 450,000-vote margin for Kennedy, and the polls showed Kennedy in the lead late in the campaign, but close enough that it could turn.

Edmund Kallina doesn't buy Hersh's evidence. Even if Nixon lost votes, he didn't lose enough to lose Illinois. And even if Kennedy had lost Illinois, he would still have won the national race. The Daley Machine, and certainly not the Giancana mob, didn't change the outcome. Kennedy didn't need help, even from Richard Daley's Machine. He took Illinois by 0.19 percent, but that is more than his margin in Hawaii (0.06), and larger than his national margin (0.17). He won Illinois by a nose while the Senate (Paul Douglas) and gubernatorial (Otto Kerner) candidates were winning easily.

Cook County went for Kennedy by 1,378,343 to 1,059,607. Kerner got 1,455,674 votes to the republican's 937,625, and Douglas got 1,407,775 to 970,050. Nixon was running up to 100,000 votes better than other republicans in Cook County, and Kennedy was lagging behind other democrats by 30,000 to 50,000. If votes were stolen for Kennedy, why did he lag so far behind the other democrats while Nixon was faring better than the other republicans? And the House delegation had 14 democrats to 11 republicans. The democrats were strong in 1960.

It's a given that voter turnout always is higher than the total vote for a given office, and normal presidential elections show a turnout of 1 to 2 percent higher than presidential votes. In 1960, Illinois voter turnout was 1.4 percent

higher than presidential votes. Cook County's turnout was notably higher, 2.2 percent. Four times fewer people voted for president in the county than in the rest of Illinois, a sign that democrats didn't like Kennedy but didn't like Nixon either. With 60,000 voters who didn't vote for president, there were ample opportunities to manipulate the vote to give Kennedy 10,000 to 20,000 more votes without being conspicuous.

Even if Kennedy had lost Illinois and Hawaii, he would still have won with 273 electoral votes, four more than enough. If Missouri had been taken, the election would have gone to the House for resolution. There the democrats dominated 262 to 174, 26 of 50 delegations. Kennedy didn't need the mafia. He won fairly, any way you look at it.

Conclusion

There was no need to fix the election. Standard Chicago politics were sufficient to guarantee a landslide for a popular Catholic democratic candidate. Daley's system for running the city gave no room for error. The Chicago Democratic Party had a tight, centralized decision-making core but allowed wide latitude at the precinct level to interpret the decision. The central committee consisted of representatives of each of the 50 wards and 30 townships, with voting based on democratic votes. Each ward had a committeeman, commonly the alderman, who handled all patronage, jobs, and other benefits. These were elected positions, and incumbents were nearly impossible to unseat unless the central committee backed the challenger. The next level below was the precinct captain, one for each of the 3,771 Cook County precincts in 1961. The precinct captain was the face of the party on a day-to-day basis. Like the committeeman, the test of the captain was his ability to deliver the vote. If he couldn't, he could lose his position and the patronage job as well. And the party had a massive amount of patronage to hand out—one estimate says it had the equivalent of 300,000 votes. And both labor and business saw the benefit of stability and were willing to add more money to the election kitty. A ward cost $20,000 in 1960.

Having controlled for other variables, Binder concluded that "someone forgot to tell the horse," a punch line from an old joke about a "fixed" race where those with inside dope lost their shirts betting on the guaranteed outcome. Along with the questionable sources and implausible stories, the analysis of the data reveals that "someone forgot to tell the voters" that the fix was in. The evidence does not show that unionists nationally or in the mob-influenced states voted heavily for the democrats. Rather, the evidence shows that nonteamster unionists in the mob-influenced states voted against Kennedy in abnormally large numbers. In Chicago, there is nothing to show that the mob's wards voted heavily for Kennedy. In two races where the mob did put in the effort, the anti-Adamowski effort and the Daley 1955 race, the mob did put up the numbers. Any

benefit for Kennedy in 1960 was the spillover from the anti-Adamowski straight ticket voting. Further, in the count nationally or in states with strong mob influence, the evidence does not show any surge of union support for Kennedy.

If anyone had mafia aid, it was Nixon. According to Bill Bonanno, son of Joe Bonanno the New York godfather, Nixon did business through Rebozo for years with Santos Trafficante's family and made real estate profits, funding anti-Castro activity and casinos and the like. And Nixon shared Hoffa's hatred for the Kennedys. Bobby had been trying to jail Hoffa since 1956 and the anticrime hearings. Hoffa had 2 million members when he threw teamster support to Nixon in 1960, and support meant more than just a get-out-the-vote effort. Hoffa had a mafia style and structure to his organization. Hoffa was Nixon's bagman for half a million in New Orleans mafia cash and another half million in New Jersey and Florida mob money. The meeting is confirmed by a Louisiana teamster official turned FBI informant as well as by a former top FBI official William Sullivan.

William Roemer was a former FBI agent. He contended that it didn't matter whether or not Sinatra told Giancana that the Kennedys would back off. Giancana had influence over the "West Side Bloc," but Daley had more influence with this group of west-side and Loop politicians. Giancana and Daley had the same interest, but Daley had the greater clout—and Roemer had two informants in high places in Giancana's outfit. Daley might have altered the vote; Giancana probably didn't, at least not enough to matter.

If the mob had a deal, it broke it. The mob had no reason to make a deal and break it. If there was no double-cross by the mob, there was no double-cross by the Kennedys. The mob had no need to kill the Kennedys because it either ignored the election or double-crossed them and thus had no revenge motive. If there was a deal, nobody benefited. More likely, there was no deal and, thus, no double-cross. The mob did not back Kennedy, did not get extra votes for him, and did not do any favors for somebody who was clearly out to get them.

And why on earth would the mob trust the Kennedys? They'd just been attacking the mob in McClellan. Reports within a couple of days of the election had Kennedy cracking down, targeting the Chicago Outfit, based on his McClellan work. Even mobster Mickey Cohen credits the Chicago outcome to the machine, not the mob.

According to the Hersh version, McDonnell said Kennedy made the deal with Giancana, and the mob got out the vote, spent retirement funds, but didn't stuff ballot boxes. Why would Kennedy need money or be concerned enough about getting the vote out (in Daley's Chicago) that he would meet with a notorious mobster, Sam Giancana? The Joe Kennedy–Sam Giancana meeting makes no sense because there was nothing to be gained. And consider the costs. Would Kennedy have risked being spotted or overheard? That would have been disastrous to the campaign, a firm link with organized crime.

The sources used by Hersh include unpublished memoirs, FBI files newly opened, as well as previously unheard Kennedy tapes and interviews. The interviews are from secondary sources—children, widows, and mistresses of Sinatra, for example—and low level associates and relatives of higher-level figures in the mob, the Giancanas for instance. And Hersh had to edit his draft work to take out citations to the "Cusack papers" after experts determined that they were fakes. Hersh kept in all of the sex, all of the Camelot debunking, and the 1960 Joe Kennedy–Sam Giancana meeting to get the mob-run unions to use their money and cash in return for a friendlier attitude by the presidency toward the mob. He should have edited more tightly and weighed the evidence more seriously. But, then, that would have been a different book, wouldn't it?

References and Further Reading

Bearak, Barry. "Book Portrays J.F.K. as Reckless and Immoral." *New York Times* (November 9, 1997).

Binder, John J. "Organized Crime and the 1960 Presidential Election." *Public Choice* (March 2007): 251–66.

Brashler, William. *The Don: The Life and Death of Sam Giancana*. New York: Harper and Row, 1977.

Exner, Judith Campbell. *My Story* (as told to Ovid Demaris). New York: Grove, 1977.

Fulsom, Don. "The Mob's President: Richard Nixon's Secret Ties to the Mafia." *Crime Magazine* (February 5, 2006).

Giancana, Sam, and Chuck Giancana. *Double Cross: The Explosive Inside Story of the Mobster Who Controlled America*. New York: Warner, 1992.

Greenberg, David. "Was Nixon Robbed?" *Slate* (October 16, 2000).

Hersh, Seymour. *The Dark Side of Camelot*. New York: Little, Brown, 1997.

Kallina, Edmund. *Courthouse over White House: Chicago and the Presidential Election of 1960*. Tallahassee: University Press of Florida, 1988.

Kelly, Kitty. "The Dark Side of Camelot." *People* (February 29, 1988).

Obrien, Michael. *John F. Kennedy: A Biography*. New York: Macmillan, 2005.

O'Connor, Len. *Clout: Mayor Daley and His City*. Chicago: Henry Regnery, 1975.

Roemer, Jr., William F. *Roemer: Man against the Mob*. New York: D.I. Fine, 1989.

Russo, Gus. *The Outfit*. New York: Bloomsbury, 2001.

14

Lee Harvey Oswald was not the sole assassin of John F. Kennedy.

PRO	Rajshekhar
CON	Tim J. Watts

PRO

More than four decades have passed since the assassination of one of America's greatest presidents. On November 22, 1963, at around 12:30 P.M.—John F. Kennedy (1917–1963), the 35th U.S. president, was shot by Lee Harvey Oswald (1939–1963) from the sixth-floor window of the Texas Schoolbook Depository in Dallas, Texas. But many American people, scholars, and researchers haven't accepted the official version of the Warren Commission report that Oswald was the sole assassin of Kennedy. The Warren Commission findings, 10 months after the national tragedy, were that no evidence was found of any connection between the crime of Lee Harvey Oswald and the city's general atmosphere of hate. But President Lyndon B. Johnson's statement, seven and a half months after the Warren Commission, reported that "the Attorney General had agreed with this interpretation and had 'discussed the practical problems at hand—problems of special urgency because we did not at that time have any information as to the motivation of the assassination or its possible implications" suggests major shortcomings in the investigation and indicates a possible conspiracy that Lee Harvey Oswald did not act alone and that the report was submitted in haste because the nation wanted an immediate answer from the government concerning the death of their beloved president (Manchester 1967: 383–84). Even the younger brother of the deceased president, Attorney General Robert Kennedy, was surprised and unresponsive to Johnson's statement, although he was not among those who suspected a grand conspiracy yet he was unclear what Johnson meant (Manchester 1967: 383–94).

Moreover, from the winter of 1963 to 1964, the mysterious deaths of many who had been related, even in the periphery, to the events of Kennedy's assassination further suggested that there was a great conspiracy to eliminate President Kennedy. Warren Reynolds, a used car lot employee who had witnessed Lee Oswald's flight after the shooting of J. D. Tippit, was himself shot in the same city on the evening of January 23, 1964. The rifleman in that shooting was seen but never found; one man was picked up but released on the testimony of a woman, who after her subsequent arrest on charges of disorderly conduct hanged

309

herself in a Dallas cell. The general who had welcomed Kennedy to San Antonio on behalf of the air force, the waiter who had served his last breakfast in Fort Worth, and the advertising director of the *Dallas News* were all found dead soon after the assassination. The advertising director, who was 45 years old, had been in excellent health. Two years later, Earlene Roberts, Oswald's landlady, died of a stroke, and Bill Whaley, his taxi driver, was killed in a traffic accident (Manchester 1967: 379). Interestingly, police officer J. D. Tippit, who had been following Oswald 20 minutes before Vice President Lyndon B. Johnson boarded the plane from Park Land Hospital (where Kennedy was taken for treatment after he was shot) on the same day had been shot to death. And when Mrs. Tippit inquired about him at 1.56 P.M. after the incident, the Dallas police radio states an incorrect home address for J. D. Tippit, who was an active member of the enforcement agency, all indicating a grand conspiracy that Oswald did not act alone (Manchester 1967: 391). And if any evidence exists of conspiracy beyond Oswald, then Oswald surely was not alone in killing Kennedy.

Vice President Lyndon B. Johnson

It is also one of the tragedies of modern U.S. history that, apart from conspiracy of international communists' and mafia groups' involvement, people have also strongly suggested that perhaps then Vice President Lyndon B. Johnson may have been involved to eliminate Kennedy to open the door to himself becoming president. Those who believe in Johnson's involvement in the conspiracy may have based their feelings on emotions at that time. But it is interesting that Johnson, a veteran democratic leader and senior to Kennedy in politics, came from Texas and was unsuccessfully contested in the July 1960 Los Angles Democratic Convention for the presidency against Kennedy. It is also true that Johnson was not the first choice of Robert Kennedy, John F. Kennedy, and his supporters for the vice presidency. Hubert Humphrey, Stuart Symington, or Henry Jackson had been expecting the nomination for vice president in July 1960 at the Los Angles Democratic Convention. In addition, when Johnson was chosen by Kennedy as his vice president to unite all democrats before the upcoming election, the political mentor and proposer of Johnson's candidacy for president Sam Rayburn (speaker of the Congress) said to Kennedy, "well, up until thirty minutes ago I was against it, and I have withheld a final decision until I could really find out what was in your heart . . . *there is always the thought in a fellow's [LBJ] mind that he might get to be President*, Lyndon is a good soldier, and he will hear the call of duty. I yield on one condition . . . that you go on the radio or television and tell the people you came to us and asked for this thing," and Kennedy agreed (Schlesinger 1967: 50–54, emphasis added).

It is interesting that in Johnson's political career the chance to get elected into Congress came to him after the death of the representative from central Texas—James P. Buchanan in 1937—and the chance to become president of the United States came after the death of Kennedy in 1963. Former Vice President Richard M. Nixon, with whom Johnson was close, stated "Johnson was one of the ablest political craftsmen of our times," and from a Senate democrat, regarding Johnson's personality, "Lyndon gives and takes. If you go along with him, he gives you a little here and there. *Lyndon never forgets*" (Mooney 1964: 27, emphasis added).

Moreover, the assassination took place in Dallas, the home turf of Johnson's political career. Apart from Lyndon, his father, Sam Early Johnson, Jr., served five terms in the Texas legislature. His maternal grandfather also saw service in the legislative branch as well as in the office of the Texas secretary of state. His maternal grandmother was the niece of a man who signed the Texas Declaration of Independence from Mexico, fought in the freedom-winning Battle of San Jacinto, and became a member of the First Congress of the Republic of Texas (Mooney 1964: 29). During the Franklin D. Roosevelt administration, in August 1935, when Johnson was appointed as state administrator for Texas of the National Youth Administration (NYA) to carry forward the New Deal's policy, he helped many idle youngsters living life without hope get jobs in Texas. He was involved in changing the lives of around 30,000 young men and women through NYA, of whom 18,000 were given assistance in getting through high school and college and others obtained jobs through the NYA. And soon his work rewarded him with the promotion as the national director of the NYA under the Roosevelt administration, with the opportunity to help more unemployed youth get government jobs. Johnson himself acknowledges in his complete authorized illustrated biography by Booth Mooney that some of these unemployed youths were in jobs and had become friends of his and played an important role in his first congressional election in 1937 (Mooney 1964: 45). Even during his senatorial election in 1948 for Texas against Coke Stevenson (former lieutenant governor and speaker of the Texas House of Representatives), Johnson, who lacked any ready-made organization, was helped tremendously by thousands of former NYA participants in every part of the state (Mooney 1964: 71). Moreover, when Adlai Stevenson contested the presidential election against Eisenhower in 1952, Johnson's political mentor Sam Rayburn set up a special Stevenson-Sparkman campaign committee with headquarters in Dallas, and many prominent democratic officeholders conspicuously stayed away from Rayburn and his committee, but Johnson stayed with his political guru at Dallas in the fall of 1952, managing the affairs of his campaign, as he was still close with many from the Dallas youth population of 1937 (Mooney 1964: 95).

How many of his young friends, who had been closely associated with him in 1937, 1948, and 1952 campaigns, had been involved in the assassination of Kennedy is a hypothetical conjecture, but later, Oswald's assassin Jack Ruby wrote a letter from jail and states:

[Y]ou must believe me that I know what is taking place, so please with all my heart, you must believe me, because I am counting on you to save this country a lot of blood-shed. As soon as you get out you must read how Texas looks at Lyndon and it may open your eyes to a lot of things. This man is a Nazi in the worst order.

Further Ruby writes:

[I]sn't it strange that Oswald who hasn't worked a lick most of his life, should be fortunate enough to get a job at the Book Building two weeks before the president himself didn't know as to when he was to visit Dallas, now where would a jerk like Oswald get the information that the president was coming to Dallas? Only one person could have had that information, and that man was Johnson who knew weeks in advance as to what was going to happen, because he is the one who was going to arrange the trip for the president, this had been planned long before the president himself knew about, so you can figure that one out. The only one who gained by the shooting of the president was Johnson, and he was in a car in the rear and safe when the shooting took place. What would the Russians, Castro or anyone else have to gain by eliminating the president? If Johnson was so heartbroken over Kennedy, why didn't he do something for Robert Kennedy? All he did was snub him. (Ramparts 1967: 59)

Many raised doubts about Johnson's involvement in the planning of Kennedy's assassination. But later, Jack Ruby's denial of the facts before the Warren Commission and his conflicting statements along with William R. Beaver's testimony after the Ruby's polygraph tests that showed he was "psychotic depressive" cleared doubts about Johnson's involvement. However, it is interesting to note that even in his polygraph test, Jack Ruby, who went with the version of Warren Commission findings, didn't respond yes or no to two very important questions: Do you think members of your family are now in danger because of what you did? Is Mr. Fowler in danger because he is defending you? (per Clayton Fowler, chief defense consul of Ruby). If Ruby could only answer the questions in the polygraph test as yes or no, then why did he remain silent after two of the most important questions concerning the danger looming over his family and that of Fowler, which he could have easily answered? It is obvious that since he hadn't killed Kennedy, the danger to him, his family, and Fowler would not have been from Kennedy supporters, who would rather have sympathized with him for killing the known assassin of Kennedy. The danger was obviously from within the U.S. administration, which baffled Ruby as he mentioned earlier to the press before the Warren Commission interrogation.

Jack Ruby considered himself a democrat because his brother Hyman had been active in democratic ward politics in Chicago. But the Warren Commission's findings, even after the recovery of political cards from his apartment urging the election of the conservative democratic slate, stated that "no evidence which could suggest that Ruby had distributed these political literature and campaigned for any political candidates," and came to conclusion that "None of his friends or associates expressed any knowledge that he belonged to any groups interested in political issues, nor did they remember that he had discussed political problems except on rare occasions" (Salisbury 1964: 341), suggesting suppression of facts of his political affiliation and conspiracy.

In addition, Federal Bureau of Investigation (FBI) files of 1947 explicitly establish, with the heading "This is sensitive," that Jack Ruby is closely connected with Richard M. Nixon and recommends that "one Jack Rubenstein of Chicago" should not be called to testify for the Committee on Un-American Activities, for he is working for Congressman Richard M. Nixon," however, according to the Warren Commission, Ruby had no connections with Oswald, organized crime, or the government. Moreover, Nixon was in Dallas on November 20–21, 1963, one day before the assassination, and Vice President Lyndon B. Johnson had also visited Dallas six months before the assassination, on April 23, 1963. Johnson's close affinity with Nixon is a well-known fact.

Certain facts about Ruby clearly appear to have been suppressed, proving that conspiracy existed, but whether Johnson was involved in it is still pure conjecture.

Whether or not Oswald was involved, it can't be denied that a conspiracy was at work. Multiple locations of origin for the gunfire (the depository building, the "grassy knoll," and others) should be enough to prove the theory. But even further, the House Select Committee on Assassinations reported as much in 1979. Studies by scholars such as Anthony Summers and Henry Hurt also contend that Oswald could not have acted alone. Even the remaining Kennedys, especially Jacqueline (his wife) and Robert (his brother), thought as much (Fursenko and Naftali 1998).

Recent Bullet Analysis Tests Proves More Than One Assassin

Recent research by former FBI lab metallurgist William A. Tobin confirmed that work conducted by Texas A&M University researchers Cliff Spiegelman and William D. James in 2007, published in the *Annals of Applied Statistics*, concluded that "evidence used to rule out a second assassin is fundamentally flawed and challenges the bullet analysis used by the government to conclude that Lee Harvey Oswald alone shot the two bullets that struck and killed President John F. Kennedy" (Solomon 2007: A03). Earlier, University of California–Irvine chemist Vincent P. Guinn concluded before the 1976 House Select Committee on Assassinations through bullet lead analysis that the five bullet fragments recovered from the Kennedy assassination scene came from just two bullets, which were traced to the same batch of bullets Oswald owned. At that time, Guinn's conclusions were consistent with the

1960s Warren Commission report that found Oswald had acted alone. However, the House Assassinations Committee concluded that Oswald was part of a conspiracy and that it was possible a second shooter fired one shot that missed the president.

But Tobin and James and Spiegelman refuted the claim of Guinn's theory, bought the same brand and lot of bullets used by Oswald (as the bullets from the batch are still in the market as collector's items), and analyzed their lead using new scientific methods and explained that "This finding means that the bullet fragments from the assassination that match could have come from three or more separate bullets, if the assassination fragments are derived from three or more separate bullets, then a second assassin is likely. If the five fragments came from three or more bullets that would mean a second gunman's bullet would have had to strike the president" (quoted from Solomon 2007: A03).

Science Daily took the theory even further, performing tests to see if it held up:

> Using new compositional analysis techniques not available in the 1960s, the team found that the bullet fragments involved in the assassination are not nearly as rare as previously reported. In addition, their findings show that one of the 10 test bullets from one box analyzed is considered a match to one or more of the five existing assassination fragments, meaning that the matching fragments could have come from three or more separate bullets and, therefore, more than one shooter. ("Bullet Evidence" 2007)

Other Evidence of a Conspiracy

Apart from Abraham Zapruder's short film on Kennedy's assassination, other photographs prove that there was more than one shooter. In any crime, photographs taken at the scene are one of the highest forms of evidence. In one example, a woman standing just to the left of the presidential limousine was taking photographs just as the shots were fired, including the front of the Texas Book Depository, but her film was confiscated by FBI agents. Later, the FBI refused to publish what could be the most reliable piece of evidence in the whole case, which also indicates suppression of conspiracy. The unknown woman photographer, nicknamed by researchers as the Babushka Lady, has still not been identified, which also raises doubts on the FBI's roles in the Kennedy investigation.

However, the House Select Committee on Assassinations (September 1976 to January 1979) finally accepted conspiracy as a motive in Kennedy's killing, proving Oswald was not alone, and concluded that (a) Kennedy was probably killed as a result of a conspiracy; (b) that four shots, not three, were fired during the assassination; (c) that one shot was fired from the grassy knoll, missing both Kennedy and the limousine; and (d) impulses caused by four gunshots were recorded on a police Dictaphone recording that was made by a patrolman's

Surrounded by detectives, Lee Harvey Oswald talks to the press as he is led down a corridor of the Dallas police station for another round of questioning in connection with the assassination of President John F. Kennedy, November 23, 1963. (AP Photo)

microphone that was stuck in the "on" position as his motorcycle rode near and through Dealey Plaza during and after the shooting. The committee's acoustical scientists concluded that an analysis of the recording revealed, to a degree of certainty of 95 percent or better, that one of the shots could be traced back to the grassy knoll. It further found that (e) Jack Ruby had significant ties to organized crime (earlier Warren Commission denied it); (f) that Ruby's killing of Oswald was not a spontaneous act but had the appearance of a hit designed to silence Oswald; (g) that in the months leading up to the assassination, Ruby had made long-distance phone calls to organized crime contacts, and some of these phone calls did not appear to have a viable innocent explanation; (h) that Ruby probably did not enter the Dallas Police Department basement via the Main Street ramp and might have gained access to the basement by help from someone on the police force, and that Ruby lied to the Warren Commission about the number and nature of his trips to Cuba prior to the assassination; (i) that the FBI and the Central Intelligence Agency (CIA) were deficient in supplying the commission with information in their possession that related to the assassination; (j) that the Warren Commission failed to adequately investigate the possibility of conspiracy; (k) that the pathologists who performed Kennedy's autopsy failed to perform a proper medical-legal autopsy; (l) that the security arrangements for the Dallas motorcade may have been uniquely insecure,

all proving a grand design of conspiracy to silence the truth related with Kennedy's assassination and the fact that Oswald did not act alone.

Other Warren Commission's Shortcomings in Revealing Conspiracy

Furthermore, there are many shortcomings in the findings of the Warren Commission itself, indicating that Oswald was not alone. Judgment in the case was swift, with the public being told that Oswald was the sole assassin, even before any attempt at investigation had ensued. No attempts to locate any other would-be assassins were made. Doctors were not permitted to discuss the case. The press was told that Kennedy was shot from behind when there were witnesses who claimed that the shots came from in front (Russell 1964).

Ronald Fischer and Robert Edwards, who testified before the Warren Commission, were less than convincing in their assertions that the man they saw was, indeed, Oswald. Dallas police allegedly showed Fischer a photograph of Oswald and asked if it was the man he saw, to which Fischer replied "that could have been the man" but that he was unsure (Kempton 1964).

It did not help the theory that Oswald worked alone that just prior to the 1964 presidential election, Johnson ordered the Warren Commission's documents sealed for 75 years. This was lifted in 1992, when Congress passed the President John F. Kennedy Assassination Records Collection Act. The records opened up by the act included more than 3 percent of all Warren Commission documents, more than 21 percent of the House Select Committee on

---

Clay Shaw and the Conspiracy to Assassinate Kennedy

The only person ever prosecuted over involvement with the assassination of John F. Kennedy was Clay Shaw, a prominent New Orleans businessman.

Clay Shaw was born on March 17, 1913, at Kentwood, Louisiana, the son of Gloria Shaw, an undercover agent for the prohibition department. As a teenager, he worked as a telephone operator. In World War II, he served in the U.S. army, rising to the rank of major, gaining an honorable discharge in 1946. He then started operating the International Trade Mart in New Orleans, which was involved in the sale of local and imported goods, as well as being a director of the Swiss company Permindex, later itself accused of involvement in assassination attempts on French President Charles de Gaulle. This was, it has been alleged, how Shaw became involved with right-wing paramilitary groups in Europe, including Italian fascists. In New Orleans, Shaw was best known for his efforts to protect historic buildings in the French Quarter of that city.

The case against Clay Shaw was brought by Jim Garrison, the event becoming the subject of Garrison's book *On the Trial of the Assassins* (1992) and the film *JFK* (1991). The jury took less than an hour to acquit Shaw. He remained in New Orleans and died from cancer on August 15, 1974.

Assassinations documents, and an undeterminable percentage of CIA, FBI, secret service, National Security Agency, State Department, U.S. Marine Corps, Naval Investigative Service, Defense Investigative Service, and many other U.S. government documents, to come out for public examination to explore the purpose behind Kennedy's killing, despite the proven fact that it was a conspiracy to eliminate Kennedy and that Oswald was not alone in killing the president.

Conclusion

Whether or not Lyndon B. Johnson was involved in the assassination, the death of John F. Kennedy can't be denied as part of a conspiracy emanating from U.S. sociopolitical conflicting milieu, as once referred to by both Robert and Jacqueline Kennedy, and rightly pointed out by historians Aleksandr Fursenko and Timothy Naftali as well as other scholars, including findings of the House Select Committee on Assassinations in September 1976. And if there was a conspiracy to kill the president, then Lee Harvey Oswald was not alone, even if he, indeed, killed the president. Kennedy, who once said in 1958 that "politics, in short has become one of our most neglected, our most abused, and our most ignored professions," and further quoting Walter Lippmann about politicians, "who are insecure and intimidated men," who "advance politically only as they placate, appease, bribe, seduce, bamboozle, or otherwise manage to manipulate" (Kennedy 1961: 229) himself fell victim to manipulation and intimidation five years later at Dallas in November 1963.

One of the greatest leaders of America, Kennedy had a brief premonition about his assassination. Before his last motorcade journey, at the breakfast meeting with the Fort Worth Chamber of Commerce at Texas, he said that "No one expects that our life will be easy . . . history will not permit it. . . . But we are still the keystone in the arch of freedom, and I think we will continue to do, as we have done in our past, our duty" (Schlesinger 1967:1024). In addition, while chatting with Jacqueline Kennedy and Kenneth O'Donnell at the Texas Hotel after the meeting about the role of the secret service in protecting the president, O'Donnell said that all they could do was to protect a president from unruly or overexcited crowds. But he further said that "if someone really wanted to kill a President, it was not too difficult; put a man on a high building with a telescopic rifle, and there was nothing anybody could do to defend a President's life" (Mooney 1964: 78–82). However, Kennedy, who regarded assassination as a risk inherent in a democracy, before his last trip continued his journey to fulfill the dream of "a nation which has forgotten the quality of courage which in the past has been brought to public life is not as likely to insist upon a reward that quality in its chosen leaders today of which one has forgotten" (Kennedy 1956: 1).

References and Further Reading

"Bullet Evidence Challenges Findings in JFK Assassination." *Science Daily* (May 17, 2007).

Fursenko, Aleksandr, and Timothy Naftali. "*One Hell of a Gamble*": Khrushchev, Castro and Kennedy, 1958–1964. New York: Norton, 1998.

Kempton, Murray. "Warren Report: Case for the Prosecution." *New Republic* (October 10, 1964).

Kennedy, John F. *Profiles in Courage*. New York: Harper, 1956.

Kennedy, John F. *The Strategy of Peace*. New York: Harper, 1961.

Manchester, William. *The Death of a President*. London: Pan Books, 1967.

Mooney, Booth. *The Lyndon Johnson Story*. New York: Avon Books, 1964.

Morley, Jefferson. "Conspiracy Theories." *Washington Post* (November 27, 2005).

"Oswald Not Alone in JFK Murder." *Athens Observer* (November 25, 1987): 8A.

Ramparts Magazine. *In the Shadow of Dallas: A Primer on the Assassination of President Kennedy*. San Francisco: Ramparts Magazine, 1967.

Russell, Bertrand. "16 Questions on the Assassination." *Minority of One* (September 6, 1964): 6–8.

Salisbury, Harrison E. *Report of the Warren Commission on the Assassination of President Kennedy*. New York: Bantam, 1964.

Schlesinger, Jr., Arthur M. *A Thousand Days: John F. Kennedy in the White House*. London: Mayflower Books, 1967.

Solomon, John. "Scientists Cast Doubt on Kennedy Bullet Analysis." *Washington Post* (May 17, 2007).

CON

Perhaps no other event in history has caused so much to be written as the assassination of President John F. Kennedy. The image of the young and popular president, riding in triumph through the streets of Dallas on November 22, 1963, only to be brutally murdered, is fixed in the public imagination. President Kennedy remains the president who held so much promise, whose term was cut short by someone.

The blame for Kennedy's assassination was initially and officially placed on Lee Harvey Oswald, a loner and misfit. Almost from the beginning, however, some disagreed. The concept that the most powerful man in the world could be killed by someone as strange as Oswald, working alone, was more than they could bear. Oswald's murder on November 24 by Jack Ruby prevented a trial that might have provided answers to the many nagging questions. Oswald never had the opportunity to defend himself or to justify the actions he took. The

silencing of Oswald opened the door for many who believed in a wide-reaching conspiracy. In the 40-odd years since Kennedy's assassination, all kinds of conspiracies have been proposed. These proposals range from Oswald as an active participant in a plot by the Central Intelligence Agency (CIA), Cubans, or the mafia, to Oswald as a patsy set up by others to take the blame.

The truth is simpler. Oswald acted alone to kill President Kennedy. Credible evidence points to him as the shooter, acting for his own twisted reasons. Oswald's life before November 1963 marked him as a true loner and failure at everything he attempted. He was completely self-centered and showed little ability or desire to work with others. He wanted to be someone who was remembered and respected. The idea of Oswald participating in a conspiracy to murder the president is not credible. His documented activities in the months before the assassination demonstrate his growing violence and irrational behavior.

The scientific evidence also indicates that Oswald fired three shots from the sixth floor of the Texas Schoolbook Depository. The ballistics show the path of the bullets and also show that the wounds inflicted were consistent with being fired by Oswald. No credible evidence exists of another gunman or additional shots being fired that day. Claims of a gunman on the grassy knoll or the sewer drains or anywhere else in Dealey Plaza have not been supported.

This section will describe how Oswald's life and psychological profile before the assassination make it unlikely he was part of a far-reaching conspiracy. It will also discuss the most common theories of a conspiracy involving Oswald and the evidence that discounts the existence of such conspiracies.

The Investigations

A week after the assassination, President Lyndon Johnson issued an executive order establishing a commission to investigate. The chair was Supreme Court Chief Justice Earl Warren. Known as the Warren Commission, the group issued its findings on September 24, 1964. Although the commission's analysis was faulty in some areas, it interviewed over 500 witnesses and accepted evidence from federal agencies. It did not have access to all existing evidence, but found that Oswald acted alone and not as part of a conspiracy.

Fourteen years later, theories about conspiracies led the House of Representatives to form a committee to review the assassinations of Kennedy and Martin Luther King, Jr. The committee's report revealed much evidence that had been withheld from the Warren Commission. It was ready to report that Oswald had acted alone, when flawed acoustical evidence was introduced. The committee's findings were amended. The final report speculated on whether organized crime had been involved in the assassination. No credible evidence connecting Oswald with the Mafia was produced. Much of the solid evidence against a conspiracy was overshadowed by the committee's final report.

Chief Justice Earl Warren

Earl Warren had been appointed to the U.S. Supreme Court by Dwight D. Eisenhower, who wanted a conservative justice in the Court. Warren, from a Scandinavian migrant family, had been born in 1891 Los Angeles and studied law, working in legal firms and then in the U.S. army before becoming a district attorney. He later entered state politics.

In 1938 Warren had been elected attorney general of California, having won the nomination of the Democratic Party, the Republican Party, and the Progressive Party. During his term in office he was involved in the internment of Japanese Americans. In 1943 he became governor of California, the only person to have been elected to that position three times, and in 1952 he was nominated by the republicans for the U.S. presidency.

After his appointment in 1953 to the U.S. Supreme Court, he was involved in a large number of liberal decisions, which angered Eisenhower. Anticommunists from the John Birch Society denounced him and urged for his impeachment. He retired from the Supreme Court in 1969, five months after swearing in Richard Nixon as president. He died in 1974.

Conspiracy Theories

A Gallup poll taken on November 29, 1963, revealed that only 29 percent of Americans believed Oswald acted alone. Hundreds of books and articles outlining conspiracy theories soon appeared. Claims by many people to know something about plots connected with the assassination helped to undermine the public confidence in the validity of the Warren Commission's final report. Although some were discredited, revelations that some agencies, especially the CIA, had withheld evidence from the Warren Commission encouraged ordinary people to launch their own investigations.

The only prosecution for the assassination was launched in November 1966 by District Attorney Jim Garrison of New Orleans. Garrison was a flamboyant prosecutor who had a history of headlines-grabbing statements that were not followed up by prosecutions. He claimed that businessman Clay Shaw had been involved in a conspiracy to kill Kennedy, along with Oswald and others. Garrison used the media to claim an increasingly large conspiracy. Despite evidence of witness tampering, Garrison's prosecution got under way in 1969. After two months of testimony, the jury found Garrison's prosecution was without merit. Shaw died before he could sue Garrison for abuse of process. Garrison later wrote about his experiences.

The findings of the House Committee in 1978 failed to convince many that only Oswald was responsible. In 1992 Oliver Stone's movie about the assassination *JFK* came out. In it, Stone included most of the conspiracy theories.

As a result, Americans who were not alive when Kennedy was killed became convinced that Oswald did not act alone.

Oswald's Life and Its Effect on His Personality

Oswald had a difficult and unstable childhood. As a result he became a self-centered, poorly educated, and badly adjusted adult. Oswald's mother spoiled him by all accounts. His father died before Oswald was born, but his mother married again when he was five. That marriage quickly became troubled and ended in a divorce in 1948. Oswald looked up to his older brother and half-brother, but lacked other male figures in his life. His mother was an unstable but dominating personality. She and Oswald moved 21 times between his birth and when he left home. Oswald attended 12 different schools before dropping out of high school. He often refused to go to school. When he did attend, Oswald was troublesome and picked on smaller children. His grades were poor and he kept to himself as much as possible.

In April 1953, Oswald was sentenced to three weeks in Youth House, a New York City correctional facility, because of his truancy. During that time, he was tested and underwent psychiatric evaluation. Oswald continued to be a loner and reject authority. The diagnosis by Renatus Hartogs, the staff psychiatrist, described Oswald as extremely self-centered and detached. He also showed dangerous traits with the potential for explosive and aggressive actions. Hartogs also noted that Oswald hid his feelings of shyness, anxiety, and insecurity. The New York Domestic Relations Court was moved by Hartogs's diagnosis to assign a probation officer to Oswald and to order treatment for him. Oswald's mother obstructed the treatment. She moved both of them from New York to New Orleans at the beginning of 1954, to prevent Oswald's placement in a residential care center.

During his childhood, Oswald became a self-taught Marxist. He read books and articles and became convinced that the capitalist system was flawed. The Soviet Union became Oswald's ideal society and he dreamed of moving there. He apparently believed that he would be accepted and honored for his understanding of Marxism.

Ironically for someone who espoused Marxism, Oswald joined the U.S. marine corps as soon as he could, on October 25, 1956. He apparently saw it as a way to escape from his mother's domination. In the marines, Oswald's character reasserted itself. He soon bridled against the authority of officers and noncommissioned officers. His fellow marines picked on him and called him "Ozzie Rabbit" for his timidity. During his time in the marines, Oswald qualified as a sharpshooter, the second-highest ranking in the corps. He was trained as a radar operator and arrived in Japan in September 1957. Although he was stationed at the same base as a U-2 detachment, no evidence exists that Oswald knew anything about the spy plane. While Oswald was overseas, he managed to shoot himself in

the arm while playing with a private weapon and suffered an emotional break-down. He was court-martialed twice. Although some conspiracy theorists believe Oswald was recruited while in the marines as a U.S. intelligence agent, he was constantly with other marines. These men have testified that Oswald was a poorly adjusted young man with no opportunity to undergo intelligence training.

Oswald became extremely disenchanted with the marines and openly studied Marxist literature and the Russian language. His failures to achieve recognition in school and in the service spurred him to plan to defect to the Soviet Union. He obtained a dependency discharge on September 11, 1959. He had already been accepted by a Swiss school, giving Oswald a reason to leave the United States. After arriving in Europe, Oswald applied for and received a visa to visit the Soviet Union in October 1959. While in Moscow, he told the Soviets that he wanted to defect. Soviet records reveal that they did not want Oswald. Only his attempted suicide on October 21 prevented Oswald's deportation. Soviet authorities ordered a psychiatric examination. Doctors who were not part of the KGB found Oswald mentally unstable. The Soviets decided to allow Oswald to remain in the Soviet Union only because they did not want to be viewed as mistreating an American citizen. Once again, some conspiracy supporters believe Oswald was recruited as a Soviet agent during this time. KGB records discount this possibility and indicate Soviet intelligence had little desire to be associated with Oswald.

Oswald was ordered to live in Minsk. At first, he was content because his de-fector status made him a celebrity. As the reality of life in the Soviet Union became clear, however, Oswald was increasingly disenchanted with his life. He found the Soviet Union was not the classless society he imagined. His job in a ra-dio factory was boring, and he had few sources of entertainment. His affair with coworker Ella Germann turned out badly, when Germann refused his proposal of marriage. Within several months, Oswald had married Marina Prusakova. His di-ary indicates that he did not love her and married her to punish Germann. Even before he married Marina, Oswald had contacted the American Embassy in Mos-cow, asking to return to the United States. After nearly a year of waiting, Oswald and Marina were allowed to leave the Soviet Union on June 1, 1962. The time pe-riod was slightly longer than average for others in a similar situation. The couple received a loan from the State Department for their travel, a common practice under the circumstances. After a brief stay in New York, they flew to Fort Worth.

Oswald's life to this point had been filled with disappointments and failures. All qualified personnel who examined his mental state found him self-centered, withdrawn, and mentally unstable. He showed hostility toward authority and a desire for individual recognition. At no point did Oswald indicate he could be part of a conspiracy. Theories about him working for intelligence organizations are not supported by information about his personality or those who knew him best. Some claim that Oswald was replaced by a Soviet agent during his time in the Soviet Union and that he somehow fooled his mother and brothers on his

return to America. In 1981, a panel of experts examined Oswald's exhumed body and concluded that the real Oswald was indeed buried in that grave.

Oswald's Activities between June 1962 and November 1963

Over the next 18 months, Oswald's actions became more irrational and violent. He was witnessed by others as being abusive toward his wife. He worked at a series of menial jobs, which he believed were beneath him. Oswald also became obsessed with the thrill and danger of spying and taking direct action against those with whom he disagreed. Using facilities at a graphic arts company in Dallas where he was working, Oswald created documents for a phony identity. In early 1963, he purchased a rifle and revolver through the mail, using this phony identity. Experts who later examined the false identification papers and order forms confirmed that Oswald had signed them. The documents were poor forgeries.

Oswald also began to plan an assassination attempt on General Edwin Walker, a Dallas leader of the anticommunist John Birch Society. He took photographs of Walker's house and located firing positions and escape routes from the area. Before making his attempt, he had Marina take photographs of him holding his weapons and radical newspapers. He claimed to have sent one to *The Militant*, although that photo apparently disappeared after the assassination. Conspiracy theorists have claimed the photographs were faked to implicate Oswald. The 1978 House Select Committee on Assassinations commissioned a study by photography experts to find out if this was true. Using a variety of tests not available to the Warren Commission, the experts found the photographs were genuine. Details such as a grain pattern in each photo were consistent throughout. Additionally, identifying marks on the rifle found in the Texas Schoolbook Depository were found to be identical to those on the rifle Oswald was holding in the photographs.

Oswald was fired from his job on April 1 for his failure to get along with other workers and declining quality of work. He did not tell his wife, but spent part of the next few days practicing with his rifle. On April 10, he told Marina he had been fired. Oswald left after supper without telling Marina where he was going. While he was out, Marina found a message from him, indicating that he did not expect to return. When Oswald showed up at 11:30 P.M., he was tense and out of breath. Oswald admitted to Marina that he had tried to shoot Walker.

Oswald had fired on Walker as he sat at a desk. The bullet had hit a wooden frame in the middle of the window and been deflected slightly. Instead of striking Walker in the head, it passed through his hair. Walker was slightly wounded in the right arm by fragments. The police were quickly called, but they found no evidence, and the investigation of the crime went nowhere.

At the end of April 1963, Oswald moved to New Orleans, where he had lived as a child. He contacted relatives and lived with them while looking for a job. After eventually finding one as a maintenance man, Oswald resumed his

radical activities. He tried to set up a chapter of the Fair Play for Cuba organization, although he was apparently the only member. Marina joined him in early May. Conspiracy followers claim Oswald had connections with the Federal Bureau of Investigation (FBI), CIA, and mafia while in New Orleans. He did use the address 544 Camp Street as a fake address for his chapter. The building had been the site of an anti-Castro organization in 1962, and Oswald may have selected the address in an attempt to embarrass his political opponents. Other witnesses claimed to have connected Oswald with David Ferrie and Clay Shaw. District Attorney William Garrison based his conspiracy case against Clay Shaw on these witnesses and other questionable evidence. Later review of Garrison's claims revealed that witnesses had been coached and contradictions in their stories were ironed out before the trial.

The FBI maintained a file on Oswald beginning with his return to Dallas in 1962 and his move to New Orleans. Despite conspiracy claims that this shows Oswald was an FBI informant, such a file was standard procedure. Oswald himself blamed the FBI for causing him to lose several jobs. When he moved from New Orleans to Dallas in September 1963, the FBI took several weeks to relocate him. Agent James Hosty interviewed Marina Oswald several times before the assassination, but never met with Oswald himself until afterward.

After being fired from his job in New Orleans, Oswald sent his wife to live in Dallas with her friend Ruth Paine. Oswald himself traveled to Mexico City at the end of September to try to get visas to Cuba and the Soviet Union. He was crushed when both the Soviet and Cuban embassies refused his requests. Once again, those who believe in a conspiracy claim that this was an Oswald imposter. The witnesses who claim to have seen Oswald at the same time he was supposed to be in Mexico City fail to be credible or describe a man who does not fit Oswald's description.

When Oswald returned to Dallas, he tried hard to get a job. On October 17, 1963, he began working as a clerk at the Texas Schoolbook Depository. Paine and some neighbors helped him get the job. Although some believe this was more than a coincidence, Oswald's job at the depository would have required a large conspiracy, including a number of Dallas housewives. At the time he was hired, the motorcade route for President Kennedy had not been set, making it impossible for his job to be part of a conspiracy.

The Assassination

On November 19, the route of the motorcade to take Kennedy to the Trade Mart was printed in the Dallas newspapers. Oswald would have seen the route and realized that it would take the president immediately in front of the depository. He had been staying at a rooming house during the week and visiting his wife only on weekends. On Thursday, November 21, Oswald broke his pattern and

caught a ride to Paine's house. He tried to smooth the strained relationship with his wife, but failed. The next morning, he returned to Dallas with a brown paper package he said contained curtain rods.

Oswald was seen by other workers on the sixth floor of the depository shortly before noon. After they left for lunch, he was able to construct a sniper's nest with boxes of books at the southeast window. The window gave him a good view of the motorcade's route, including down Elm Street, immediately in front of the depository.

At 12:30 P.M., President Kennedy's motorcade turned on to Elm Street. Three shots were fired. Kennedy was hit by one shot in the back, with the bullet exiting out his neck. The last shot struck him in the head and blew off much of the right side of his head. Governor John Connally, riding just ahead of Kennedy, was hit in the back and wrist. The motorcade immediately raced to Parkland Hospital, where Kennedy was pronounced dead at 1:00 P.M.

When police surrounded the depository, Oswald was not in the building. He left and proceeded by bus and cab to his rooming house. Oswald changed his shirt and left. About 1:15 P.M., Officer J. D. Tippit pulled his car up behind the walking Oswald and called to him. He fit the description that had been broadcast of the possible assassin. Oswald exchanged some words with Tippit, who then got out of his car. As he rounded the front, Oswald pulled out his revolver

President John F. Kennedy, Governor John Connally of Texas, and First Lady Jacqueline Kennedy ride through Dallas, Texas on November 22, 1963, moments before the president was killed by an assassin. (Library of Congress)

and shot Tippit dead. Eyewitnesses used Tippit's radio to call for help. Others trailed Oswald to a nearby movie theater. At 1:46 P.M., police officers stormed into the theater. When they approached Oswald, he hit the first policeman and pulled his revolver. The gun failed to fire and Oswald was subdued.

The next 48 hours at police headquarters were a media circus, as Oswald was paraded before the press between interrogations. Oswald denied any guilt, while observers felt he was enjoying the attention he received. Howard Brennan, who saw Oswald firing from the sixth floor window, and witnesses to Tippit's murder picked Oswald out of lineups. He was indicted for the murders of Kennedy and Tippit. On Saturday, November 23, the police announced Oswald would be transferred to the sheriff's jail. The transfer was delayed until Sunday, November 24. As Oswald was being led to a waiting car, Jack Ruby ran out of the crowd and shot him once. Oswald died shortly afterward at Parkland Hospital. He made no statements after being shot.

Specific Conspiracy Charges: How Many Shots Were Fired?

Most of the witnesses in Dealey Plaza agreed that three shots were fired during the assassination. Three men on the fifth floor of the depository heard three shots fired directly above them. At least one man reported to police immediately afterward that he could hear the sounds of a bolt-action rifle being worked and three shell casings hitting the floor. When police searched the sixth floor, they found Oswald's rifle and three empty shell casings. The Warren Commission accepted this number of shots. Using the Zapruder film of the assassination, they established a timeline for the three shots.

Of the witnesses in Dealey Plaza on November 22, only 2 percent believed there were more than three shots fired. Generally accepted evidence indicates that three shots were fired from the depository. The House Committee on Assassinations in 1978 came to the conclusion that four shots had been fired that day, indicating a second gunman. Their conclusion was based on a Dictaphone tape from the minutes around the assassination. The microphone on a police motorcycle had been stuck in the on position. If the motorcycle had been in Dealey Plaza, the committee believed that the shots would have been recorded. The committee eventually assumed the motorcycle was in Dealey, but the presumed rider denied it was his. Experts studied the tape and believed they found evidence of four shots. In the hurry to wrap up its investigation, the House Committee accepted the findings. The conclusion that Oswald had killed Kennedy as part of a larger conspiracy was based largely on this recording.

Within months the audio evidence was discredited. On the tape, the sound of police sirens does not appear until two minutes after the presumed gunshots. If the tape was from a motorcycle in the motorcade, it would have meant that the policeman waited two minutes after the shooting, then tried to catch up with

the motorcade. Instead, it appears the recording is from a motorcycle at the Trade Mart. Even more damning, part of an order from Sheriff Bill Decker, given a minute after the shooting, can be heard at the presumed moment of the assassination. The timeline the experts proposed for the recording was incorrect. The time in which they believed they detected gunshots was at least a minute after the shots had been fired.

Did Oswald Have Time to Fire Three Shots?

The Warren Commission established that all three shots were fired within six seconds. Only two struck Kennedy. The commission believed that the first shot hit the president and Governor Connally. In the Zapruder film, the presidential limousine disappears behind a road sign. When it emerges, Kennedy had been wounded. The second shot was believed to have missed. The third shot was the fatal head shot. Using the Zapruder film to measure the time between the shots, the commission believed that Oswald had only 4.8 to 5.6 seconds to fire all three shots, a very short time in which to work a bolt-action rifle. Conspiracy supporters have used this short period to claim that Oswald could not have fired three times. Therefore, another shooter must have been in Dealey Plaza.

Tests by the Warren Commission proved it was possible to fire Oswald's rifle within the time period. Computer enhancements, however, led the House Committee in 1978 to project a different timeline. The evidence they had indicates that Oswald's first shot missed. Before Kennedy disappeared behind the sign, he and others can been seen reacting to something. Most notably, a young girl running parallel to the limousine stopped and turned toward the depository. She testified she turned because she heard a shot. The timing of the first shot would have given Oswald between 8 and 8.4 seconds from the first to the third shot, nearly twice the time the Warren Commission supposed. Various tests with riflemen unpracticed with bolt-action weapons showed they were able to get off three shots in this amount of time. When the tests were run with a moving target at the same distance as Oswald was from Kennedy, the riflemen averaged two hits out of three shots, the same as Oswald. Furthermore, Marina testified that Oswald spent many evenings in New Orleans in their apartment, just working the bolt-action of his rifle, again and again.

Evidence that one bullet had missed was found immediately after the assassination. A man standing 500 feet away, in a direct line from the depository and the president's limousine, was struck in the cheek by a tiny piece of concrete kicked up by a bullet. This appears to have been the first bullet Oswald fired.

The Single-Bullet Theory

One of the most compelling arguments that Oswald did not act alone involved the so-called single-bullet theory. The Warren Commission believed that one

bullet had hit Kennedy in the back, traveled through his neck, and exited near his collar and tie knot. The bullet then struck Connally in the right back, passed through his chest, and hit the governor's right wrist. It broke Connally's wrist before ricocheting into his left thigh. This bullet fell out of Connally's thigh at Parkland Hospital, where it was recovered. Remarkably, the bullet was only slightly deformed. Although sometimes described as "pristine," the bullet is flattened and part of its lead core is exposed. Conspiracy theorists claim a single bullet could not have been fired from the Schoolbook Depository and inflicted this damage without making a turn in flight. They especially dispute whether the bullet could have done so much harm and not been more deformed.

The answer to part of this issue depends on understanding where Connally was sitting in relationship to the president. Connally and his wife were sitting ahead of President and Mrs. Kennedy, on jump-seats. The seats were slightly inboard of the Kennedy's positions. Connally was not directly in front of Kennedy, but slightly to the left. The bullet that struck Kennedy was traveling slightly from right to left, so it hit Connally in the right shoulder after exiting Kennedy. It was deflected slightly after hitting Connally's rib. From there, the bullet traveled in a straight line. Connally was turned slightly to the right, so his right wrist and left thigh were in line with the exit wound in his chest.

There is also a question of Connally's reaction in the Zapruder film. According to some critics, he does not seem to react to being hit for up to two seconds after Kennedy has clearly been hit. The Warren Commission and the House Committee both postulated that Connally was just slow to react. Such delays are not unknown among gunshot victims, but conspiracy supporters believe it shows Connally was hit by a separate bullet. In 1992, enhancement of the Zapruder film became available for study. From the time Kennedy and bystanders reacted to the first bullet, which missed, to the time the car became visible from behind the road sign, about 3.5 seconds elapsed. This time was sufficient for Oswald to have chambered another round and prepared to fire. Kennedy is beginning to react to being hit. His arms are raised, with the elbows out and hands clenched near his throat. The reaction is known as the Thorburn's position, and it occurs when the upper spinal cord is damaged. X-rays revealed that a stress fracture to the C6 vertebrae resulted from the passage of the bullet that struck Kennedy in the back. Kennedy's movement was consistent with medical observations.

Even though Kennedy reacted, Connally did not seem to have been hit. The enhanced film, however, showed that at the same time Kennedy was hit, the right lapel of Connally's suit flipped out from his body. The movement was in the same area as the exit wound Connally suffered. Connally's physical reaction to being hit was revealed by the movement of his right hand at the same time. It moved sharply up and down. The most obvious reaction occurred when Connally's mouth opened and his cheeks puffed out. A noticeable period elapsed before this happened because it was caused by Connally taking his next breath.

He suffered a collapsed lung, and the reaction shows his body trying to draw in air. The timing in the Zapruder film shows that Connally was certainly hit by the same bullet that hit Kennedy.

The bullet that struck Kennedy was traveling about 1,700 feet per second when it struck him. It penetrated Kennedy's body without striking any bones or other hard surfaces. When it exited, it was traveling about 1,500 feet per second. The bullet started tumbling as it exited as well. Connally's entrance wound was about 1.25 inches long, showing that it struck him while traveling sideways. The bullet struck Connally's rib, which changed its path and slowed it even more. The bullet was still tumbling when it exited at 900 feet per second. It was traveling backward and struck Connally's wrist, breaking his radius. The bullet continued in a straight line through Connally's wrist and struck his left thigh at about 400 feet per second. The momentum was only sufficient to penetrate the skin.

Tests conducted by the Warren Commission failed to give the same kinds of wounds or bullets that were not badly deformed. Investigators, however, used bullets traveling at full velocity on test subjects. It was not until the 1990s that tests with reduced charges showed it was possible to cause such wounds without deformities resulting.

Conspiracy theorists also fail to note that when the bullet was found at Parkland Hospital, complete examinations of Kennedy and Connally had not taken place. If authorities later found fragments that showed three bullets had struck the men, and another, completely different, bullet was produced, it would have caused the conspiracy to unravel.

The Grassy Knoll

Many conspiracy theories claim a second gunman was located on the grassy knoll ahead and to the right of the president's limousine. Because of the echoes in Dealey Plaza, many witnesses were unable to determine where the shots came from. As the motorcade sped away, some people ran toward the grassy knoll. A fence at the top has been cited by many conspiracy supporters as the cover from which a second gunman fired on Kennedy.

In the years following the assassination, a number of people have come forward saying they saw someone or a puff of smoke from the grassy knoll. A number of them, such as Jean Hill, claimed they saw suspicious people, sometimes described as policemen on the knoll. Hill's story was disproved by pictures that showed her remaining far from the knoll until long after the motorcade had passed. Other witnesses have been discredited, such as Gordon Arnold who claimed to be there but does not appear in any of the pictures.

Other evidence, such as grainy photographs of shadows under the trees on the knoll, is cited. One in particular is claimed by conspiracy supporters as showing the outline of a man in a police uniform with a large puff of smoke

from a gun being fired. They ignore the fact that modern ammunition produces very small amounts of smoke. To explain how a shooter could fire over the fence, most explain that he stood on a car.

Other witnesses with a view of the back of the fence did report two men located near the overpass at one end, not where conspiracy supporters believe a gunman was located. These men were approached by police after the assassination. No mention was made of a weapon being seen.

Well-meaning witnesses may also have been fooled by a steam pipe that ran at the top of the knoll. Puffs of smoke could well have been escaping steam. They also failed to explain what happened to the bullet this gunman might have fired. The three bullets fired by Oswald were all accounted for. Some theorists believe the gunman fired an exploding bullet that hit Kennedy in the exact place where Oswald's third bullet exited, at the same time. Most observers agree that this would have been an extraordinary piece of marksmanship.

Did Ruby Kill Oswald to Silence Him?

Nightclub owner Jack Ruby has been linked to organized crime and a contract to kill Oswald before he could talk about a conspiracy. The evidence indicates, however, that Ruby was genuinely upset over Kennedy's death. Ruby was known as someone who would be drawn to the action. He appeared at the Dallas police headquarters on the evening of November 22. Because he was known to many policemen, Ruby was allowed to mingle with the media and approach Oswald. Photographs showing Ruby from that night indicated he was carrying his revolver with him, as he usually did. At one point, Oswald passed within two feet of Ruby, but Ruby did not act at the time.

Over the next two days, Ruby continued to grieve for Kennedy. Those who knew him felt that he was becoming more despondent. On the morning of November 24, he went to the Western Union office to wire money to an employee. He did not know that Oswald had not yet been transferred, but walked 200 feet to police headquarters. Ruby made his way down a ramp that would be used by the vehicle carrying Oswald when the policeman guarding it was distracted. He arrived in the basement only about 30 seconds before Oswald emerged.

If Ruby had been any later he would have missed the opportunity to shoot Oswald. Also, Oswald would have been transferred earlier except that an interrogation went longer than expected, and Oswald asked for a change of clothes. Any conspiracy theory that involves Ruby ignores these facts. The evidence indicates that Ruby only took advantage of the opportunity and acted out of emotional upset.

Conclusion

The assassination of the most powerful man in the world by a lone, maladjusted gunman is difficult for many people to accept. Kennedy's death becomes easier

to accept if it was the result of a wider conspiracy, acting on ideological grounds to remove him. These people want to believe that his death was not senseless or the result of an opportunistic action by Oswald. They tend to ignore evidence and even logic and look for comfort in a larger plot.

References and Further Reading

Hosty, James P., and Thomas Hosty. *Assignment, Oswald*. New York: Arcade Publishing, 1996.

Kurtz, Michael L. *The JFK Assassination Debates: Lone Gunman versus Conspiracy*. Lawrence: University Press of Kansas, 2006.

Mailer, Norman. *Oswald's Tale: An American Mystery*. New York: Random House, 1995.

Mallon, Thomas. *Mrs. Paine's Garage and the Murder of John F. Kennedy*. New York: Pantheon, 2002.

Moore, Jim. *Conspiracy of One: The Definitive Book on the Kennedy Assassination*. Fort Worth, TX: Summit Group, 1990.

Myers, Dale K. *With Malice: Lee Harvey Oswald and the Murder of Officer J. D. Tippit*. Milford, MI: Oak Cliff Press, 1998.

Nechiporenko, Oleg. *Passport to Assassination: The Never-Before-Told Story of Lee Harvey Oswald by the KGB Colonel Who Knew Him*. New York: Carol Publishing Group, 1993.

Oswald, Robert L., Myrick Land, and Barbara Land. *Lee: A Portrait of Lee Harvey Oswald*. New York: Coward-McCann, 1967.

Posner, Gerald L. *Case Closed: Lee Harvey Oswald and the Assassination of JFK*. New York: Random House, 1993.

United States Warren Commission. *Report of the President's Commission on the Assassination of President John F. Kennedy*. Washington, DC: Government Printing Office, 1964.

15

Considering the refusal of Saddam Hussein to comply with United Nations–imposed inspections, it was reasonable for George W. Bush and his advisers to assume that Iraq harbored weapons of mass destruction and that justified the invasion.

PRO   Dan Tamir

CON   Christian Nuenlist

PRO

Imagine this situation: in a somewhat tough neighborhood, a man with unknown intentions might be holding a certain weapon, the usage of which would have an unpleasant impact on his neighbors. Although there's clear evidence that this man has already used a similar weapon in the past, there is no clear, unequivocal proof that he is also holding this weapon at the moment. As a matter of fact, he claims that he has already gotten rid of it. On the other hand, he might be simply lying. If you were an active citizen in a community where there's no policeman and the city council—of which you are a member—malfunctions, what would you do?

This was, more or less, the dilemma faced by U.S. policy makers as they contemplated the 2003 invasion of Iraq. The question whether the invasion of Iraq was justified played a central role in international politics at the time and has remained highly disputed in wide circles. Unlike chronologically more remote historical debates, which usually have only an indirect influence on our life today, this controversy is still a matter not only of scientific differences among scholars, but also of bitter political disputes. Many things have changed in Iraq since the initial invasion in 2003 and during the following seven bad years of American occupation; naturally, one cannot turn back history. But as the waves of violence unleashed by the U.S. invasion and the fall of the Ba'ath regime continue to take the lives of many U.S. soldiers and many more Iraqi citizens every week, the question "Was the invasion justified?" remains relevant to the public debate, even if it is increasingly embedded in the more urging question "What are we still doing there today?"

Furthermore, besides its current relevance to contemporary U.S. and world politics, the question whether the threat of the existence of weapons of mass destruction (WMD) in the hands of Saddam Hussein justified the invasion raises some

333

other, more general questions. These questions concern the basic fundamentals of international policy and intelligence-based decision making, as practiced and experienced by different historical actors since days of yore.

Generally, the WMD dilemma comprises two distinct questions lying at its roots. The first is the age-old riddle of predictability: Can we predict future events, based on past experience? In our case, this question was manifested in the uncertainty whether Saddam Hussein's regime, which was known to have used chemical weapons before, still had more of the stuff and was willing to use it. The second question is that of risk versus chance: How much are we willing to take the risk that a wrong prediction might lead us to unwanted outcomes? In our case, the question was what we are risking if we assume he doesn't have it while he actually does.

A common belief is that U.S. invasion of Iraq was motivated primarily by the wish to secure access to one of the world's largest oil reserves. Some researchers claim there is clear evidence that the invasion of Iraq in 2003 was planned years before, as planners and decision makers became aware of the accelerating pace at which the world's oil reserves are depleting (a very simple process yet with enormous implications, usually known as "Peak Oil"). Other commentators argued that the invasion was motivated by a simple, even "primitive," private vendetta: after eight years of Bill Clinton's interregnum, the administration of George W. Bush wished to finish the job begun but not completed by his father, George H. W. Bush. Few people even still believe that behind this invasion lay the naive intention of making Iraq a modern, North Atlantic style democracy.

This section, however, will argue that, taking into account the recent history of Iraq, because of the way the United States role in the international arena was perceived by its policy makers and the intelligence available to them at the time, the invasion of Iraq was justified by the threat that ready-to-use WMD were in the hands of Saddam Hussein's regime. As a matter of fact, this justification and all the other above-mentioned motivations and reasons do not exclude one another: It might have been that the WMD threat was just one of many other true or false justifications. The search for the "smoking gun" could have proved to be futile from the very beginning, if the search is after the one unique and specific piece of evidence and not after a broader context of clues and reasons.

This section has three main parts. The first part will review the two levels—global and local—of past experience with WMD. The second part will discuss the process of intelligence work, conduct of an intelligence analysis, and show how this process might have run in the Iraqi case, leading to the assessment that Iraq had active WMD. The section will conclude with an examination of the issue in hindsight, with the attempt to conclude from it about WMD politics on the global arena today.

Other Times, Other Places: Global Precedents for the Use of WMD

The 20th century, "the Age of Extremes" as it was called by Eric Hobsbawm, saw the most horrible wars humanity has ever witnessed. These conflicts gained the infamous name of "Total Wars," as their immediate aim was the extermination of the enemy's troops, and, frequently, its civilian population as well. The weapons used in these wars are usually divided into conventional weapons (explosives, bullets, shells, and their like) and nonconventional weapons. Although both kinds of weapons can be used to kill millions, it is usually the latter ones that are called weapons of mass destruction.

WMD are munitions with a capacity to kill large numbers of humans indiscriminately. The term usually refers to biological, chemical, and atomic weapons. Biological weapons are pathogens used to spread diseases among a certain population, either military or civilian. Owing to their active and "lively" nature, it is difficult to judge when exactly in the past epidemics and diseases may have been the consequences of aimed deployment of weapons and when they were just "regular" outbreaks. It is assumed, for instance, that the great bubonic plague, devastating large parts of Europe and the Mediterranean basin in the mid-14th century (an event known as the Black Death), broke out after Tatar warriors lobbed plague-infected corpses into the besieged city of Kaffa, on the shores of the Black Sea, in 1346; from there, the disease spread in all directions.

The advancement of microbiology and immunology during the past 100 years has enabled scientists to isolate and master pathogens in laboratories, thereby permitting them to create concentrated masses of these carriers of death. The best-known example of an actual usage in battle of such weapons is probably the deployment of plague (and probably also anthrax and cholera) by Japanese forces in China during World War II. However, although also developed by some other states during the 20th century (including nazi Germany, and the United States and the Soviet Union during the Cold War), biological weapons were not widely used, mostly due to technical rather than moral reasons, as explained below.

Chemical warfare is also nothing new. Some hunter-gatherers, for instance, used vegetable and animal venoms and poisons for hunting: Arrows dipped in poison shot at game might be the earliest example of a chemical weapon in human history. As societies grew bigger and technologically advanced, so too did their application of these lethal substances: The poisoning of water systems or the fumigation of besieged fortifications were practiced by ancient Greek strategists during the Peloponnesian wars.

The modern use of chemicals as weapons of mass destruction, therefore, was less an invention and more of an innovation, differing from its older precedents more in scale than in content. One need not have a PhD in chemistry in order to imagine how suffocating smoke or burning acids blown at the enemy might reduce his ability or willingness to resist.

The birth of modern WMD, however, was connected to a couple of people who did hold a PhD in chemistry. On April 22, 1915, during the first stages of World War I, a group of German soldiers guided and supervised by Fritz Haber, the famous German chemist, exploited an eastern breeze in order to deploy chlorine on the French troops facing them. The French did not lag behind for long: soon afterward they used phosgene against the German troops. French chemical warfare was directed by another famous chemist, Victor Grignard. Both Haber and Grignard worked simultaneously on attacking with chemical weapons and developing protection from them; each of them also won the Nobel Prize in Chemistry (Grignard in 1912, Haber in 1918) for other, more benign inventions.

The usage of this new and frightening weapon caused shock and awe among citizens and soldiers alike. Lieutenant General Ferguson, commander of the 2nd British Corps at that time, was quoted as saying that "it is a cowardly form of warfare which does not commend itself to me or other English soldiers" (Cook 1999: 37). His disapproval, however, did not prevent his own army from using it too, in September of that same year. On the morning of April 17, 1917, during their attempt to break the Ottoman defense lines near Gaza, the British army first introduced chemical warfare to the area they termed "Middle East." British troops were also the first to introduce chemical weapons to Iraq, when suppressing rebels and insurgents there during the 1920s.

The American war machine also began producing lethal gases, although it did not have time to use them before the November 1918 armistice. Like every technical innovation, once the jack was out of the box, everybody wanted to have one too. The use of gas in modern warfare proliferated very quickly and was no longer confined to those European barbarians in their mutual bashing on their continent. Soon after World War I, chemical warfare was used by the Spanish military in Morocco and by Italian forces in Libya and Ethiopia.

The damage caused by chemical weapons during World War I made a considerable public impact, which led to attempts to stop or at least restrain its usage. The Geneva Protocol, which prohibits the use of gases and bacteriological methods of warfare, entered into force in February 1928, but it prohibited only the use of such weapons, not their production or storage. The Biological and Toxins Weapons Convention of 1972 did cover this aspect, but had no reliable verification mechanism. It was only the Chemical Weapons Convention, which entered into force in 1997, that has put a ban on development, production, stockpiling, and use of chemical weapons.

Mostly due to technical reasons, World War II saw less of the use of chemical weapons in the battlefield (although chemicals were widely used during the war by the nazis in their extermination camps). The end of World War II, however, saw the culmination of man's destructive capacity so far, as incarnated in the third part of WMD's unholy trinity: atomic weapons. Within four days, between August 6–9, 1945, American bombers dropped two atomic bombs on

two Japanese cities, Hiroshima and Nagasaki, destroying them to the ground. Although these were the only two deployments of atomic weapons so far, the world had entered the "atomic age." At least eight countries have acquired such weapons during the past 50 years, and some others are presumed to be developing and possessing them.

All three parts of the WMD unholy trinity carry with them the capacity of inflicting horrible damage to human beings and their environment. However, it is the chemical ones that were used most frequently, as the usage of the other two is either problematic or limited. Biological weapons are problematic because they carry with them the imminent risk of self-damaging. Accidents in their storage are the minor concern, much graver is the danger that they go out of control after deployment, ravaging not only the enemy but also the user's own troops and civilians (a danger that exists with chemicals weapons too, but to a much smaller extent). Atomic weapons, on the other hand, are limited because in order to build them one needs not only rare materials but expensive and sophisticated technologies as well. In other words, it requires a lot of knowledge, time, and money. Chemical weapons, by contrast, stand in between. Though they are neither the safest thing to hold in your bathroom closet nor the easiest thing to produce in a school lab, they are still relatively cheap and easy to deploy, if your aim is to murder as many people as possible in one go.

To sum it up, one can say that WMD, which aim to cause indiscriminate casualties and damage, became available across the board during the past century. This is a general phenomenon, where the Iraqi case is not unique. Although there is a considerable distance between *possessing* WMD and *using* them, we know that more than once, the extreme damage inflicted by these weapons has not always deterred decision makers from using them. To paraphrase Anton Chekhov's famous saying, the gun from the first act is on the wall; the question is when will it shoot. And this brings us to the specific Iraqi case.

The Gun that Already Smoked Once: Iraq's Past Use of WMD

Iraq was a party to the Geneva Protocol since 1931; it joined the Nuclear Non-Proliferation Treaty in 1969, and signed the Biological Weapons Convention in 1972 (though ratifying it only 20 years later). So officially, Iraq was a normative state, obeying international treaties and conventions. Practically, however, the Iraqi case proved that, in diplomacy like in politics, a certain disparity might occur between promise and fulfillment: today it is clear that Saddam Hussein's regime has violated at least one of these treaties—the one limiting the use of chemical weapons—if not all three.

Although some Iraqi scientists have testified they were working on building nuclear weapons during the 1980s, there is no clear evidence showing that the Iraqi government was seriously pursuing nuclear weapons in the past. It is clear,

however, that Iraq had some nuclear devices that might have been adapted and upgraded into weapons manufacturing plants. The Iraqi nuclear research program was launched in the mid-1970s. Since Iraq's local nuclear resources are limited (it probably had some uranium but of low quality, extracted from the Aqashat mines near the Syrian border), it sought materials and technology from other countries, primarily from France.

Iraqi nuclear research was carried out at seven sites. The main nuclear site was at Al Tuwaitha, about 20 km southeast from Baghdad, where in 1977 a French nuclear reactor was constructed. It was of a model called Osirak (or "Osiris"). Since Egyptian culture had only a limited popularity among Ba'ath leaders at that time, the reactor received the local name Tammuz 1 (Tammuz is the Hebrew and Arabic name for July, the month in which the Ba'ath party retook power in 1968; Tammuz was also the old Babylonian god of the summer).

It did not take long until the people in Baghdad (the "Present of God" in old Persian) were weeping for Tammuz. Surviving a failed assault by two Iranian jet fighters in September 1980 at the very beginning of the Iran-Iraq war, Tammuz was paralyzed by a second air strike on June 7, 1981, this time by Israeli bombers. The timing of the air strike was explained at that time by the reactor's activity situation: Israeli intelligence estimated that the reactor's plutonium fuel rods were soon to be shipped to Iraq from France, thus making it "hot." Damaging a "hot" reactor might have disastrous environmental consequences. The final blow to the Al Tuwaitha site was in January 1991, during the first days of Operation Desert Storm, as U.S. bombers repeatedly bombed it.

While the question is still open how close Saddam Hussein's regime was to achieving military atomic capabilities, it is quite clear that it was using chemical ones. The almost decade-long war between Iraq and Iran during the 1980s set the scene for a series of Iraqi assaults by chemical agents against enemy soldiers and civil population alike. The first report of such an attack was made in November 1980, a few months after the beginning of the war with Iran, when Tehran Radio reported an Iraqi chemical attack on Iranian forces near Susengerd. Naturally, the Iranians had an interest in blaming Iraq for whatever crime possible, trying to denigrate the Ba'ath regime in international forums and maybe even using it as a pretext for developing their own chemical weapons. The report received no corroboration from any third party at the time. But within the next few years, more and more pieces of evidence were collected, all pointing to the fact that Iraqi forces used chemical weapons to fend off Iranian infantry attacks. On March 21, 1986, the United Nations Security Council recognized that "chemical weapons on many occasions have been used by Iraqi forces against Iranian forces" (United Nations 1986: 233).

The most well-known chemical attack carried out by Iraqi forces was during the last phases of the Iran-Iraq war, in an operation that received the name Anfaal, after the Quran's eighth chapter, dealing with the conduct of war. Blatantly flouting Quranic war rules, this operation was targeted to a large extent against

civilians and noncombatants after Peshmerge, the Kurdish militia, took control of some areas of northern Iraq near the Iranian border. On March 16, 1988, military airplanes dropped bombs containing lethal chemical agents—most probably sarin, VX, and maybe also mustard gases—on the town of Halabga, located about 150 km from Baghdad. The tapes, which were leaked to the international press a few days later, clearly proved that chemical agents were used against a defenseless civil population. Saddam Hussein kept claiming for years that Iranian airplanes committed this attack; this version is very dubious, since the Iranians and their Kurdish allies held the area at the time of the attack.

The use of chemical weapons by Iraqi military did not happen in an international vacuum. Foreign governments, the United States and Great Britain included, were in no way free to pass the buck, as they did almost nothing to stop Iraqi use of chemical weapons. The United States was the only member of the UN Security Council to oppose and vote against the above-mentioned 1986 statement (while the United Kingdom, the first to bring chemical weapons to Iraq, only abstained). A few months after the well-documented attack on Halabga, the U.S. Senate unanimously passed the Prevention of Genocide Act, which put severe limitations on U.S. relations with Iraq. The Reagan administration, however, opposed the bill and eventually prevented it from taking effect; the administration saw importance in strengthening Saddam in his war against Iran.

According to some sources, U.S. governments—especially the Reagan administration—held not only indirect but also direct responsibility for the Iraqi use of chemical weapons. While doing business with them, Reagan's administration showed overt animosity toward Iran and shed no tears on seeing Iranian soldiers killed by the thousands. It is argued that during the 1980s, U.S. laboratories trained Iraqi scientists in producing biological weapons; in 1982 President Reagan removed Iraq from the list of states that sponsor terrorism, despite intelligence reports that Iraq was pursuing a biochemical warfare program.

In his book *Turmoil and Triumph* (1995), Reagan's Secretary of State George Shultz admitted that reports of Iraq using chemical weapons against Iranian troops first began "drifting in" already at the end of 1983. This did not prevent Reagan from sending on December 19, 1983, an envoy to Saddam Hussein with a letter calling to "resume our diplomatic relations with Iraq." The name of that envoy was Donald Rumsfeld.

Therefore, even if it did not help Saddam's regime in developing chemical and biological weapons, the United States at least turned a blind eye toward Iraq's development of such materials. In other words, the administration knew very well what the Iraqis had. In this light, Rumsfeld's axiom about future and past predictability is not enigmatic anymore: You don't need to predict anything when you simply know what your ex-partners have.

Last but not least, one should take into account the personal dimension of the Iraqi regime. As mentioned above, many states have used chemical weapons in the

Left: President George W. Bush declared major fighting over in Iraq, calling it "one victory in a war on terror," aboard the aircraft carrier USS *Abraham Lincoln* on May 1, 2003. Right: Iraq's president Saddam Hussein fires a rifle celebrating Iraq's commitment to liberating the holy city from Israeli rule, Baghdad, November 20, 2000. (AP Photo/J. Scott Applewhite/Jassim Mohammed)

past. However, in all these states the regimes and decision makers who ordered the deployment of WMD were ejected from power afterward. In Germany, the Reichswehr was dismantled after World War I, as was the nazi regime after World War II. The fate of the fascist regime in Italy and the military government in Japan was similar; in Britain and the United States, democratic elections replaced the parties in power sooner or later after those wars. This was not the situation in Iraq. Saddam Hussein's regime kept its power after the war with Iran. Two years after the war with Iran ended, in August 1990, the Iraqi military invaded Kuwait, beginning another round of war and bloodshed. Next to the ground invasion of Kuwaiti soil, the Iraqi military launched SS missiles on Israel and Saudi Arabia and repressed Shiite insurgencies in southern Iraq with utmost violent measures. Saddam Hussein proved, therefore, that he was sticking to his aggressive, ruthless agenda. Ruling Iraq with an iron fist, Saddam tightened his hold on the state after surviving the 1991 U.S. assault. Saddam did what dictators usually do when stressed externally: he increased domestic oppression. Knowing the hostility many Iraqis had toward him, his regime became more exclusive and centralist with the time. Although the Kurdish provinces of northern Iraq gained virtual autonomy after the 1991 war, it was only after the U.S. invasion in 2003 that his regime collapsed, bringing a final end to his regime's external aggression potential.

The Process of Intelligence Work

"War is the realm of Uncertainty," wrote Carl von Clausewitz. In order to dispel it a bit, one uses intelligence evaluations. Like all else within the military sphere, intelligence work is divided into three parts: the collection of evidence and pieces of information; the interpretation of these facts, looking for their meaning; and the evaluation of future threats and chances deriving from the situation as it is perceived. All three parts of this process are equally important; the process cannot be completed if one of the phases is omitted.

Here is an example: Paw prints in my apartment corridor are a piece of evidence for the presence of some animal there. Based on past knowledge, I may then assume and interpret it as a sign that my sister's golden retriever just came in after wallowing in a puddle outside; I can then evaluate that it is now heading toward the kitchen, with the aim of feasting on the sausages I was just about to cook there.

Intelligence analysis is a fascinating matter, but there is one thing it cannot do: It cannot tell decision makers what exactly they should do; it can merely help them understand the current situation. To get back to my example, it can help me decide what I should do next. Should I call the dog and order it out of the kitchen; try running to the kitchen in order to get there before it does; or bring a towel and clean the floor? But it cannot direct me to do any one of these actions. Any of them may be the outcome of my decision-making process, all relying on the same intelligence analysis. In other words, the intelligence provider can bring the menu, explain what each item is, and estimate how tasty it may be, but it is up to the decision maker to choose what to eat, when, and how.

Each of intelligence's three phases has its own typical problems and difficulties. In the first part, the collection of facts and evidence, the problem lies with the facts themselves. Sometimes evidence is clear and easy to get, but this is not always the case. One kind of difficulty is when evidence is blurred, partial, or incomplete; another is when items of evidence abound and contradict one another, making it hard to filter the correct, important evidence from the incorrect and nonimportant items. To all of this one should add the impact of counter-intelligence (i.e., deliberate efforts by the researched object to deceive and mislead the intelligence of the enemy by implanting false facts and disseminating wrong information).

The main problem with the second part, the interpretation, is that it is based on the past experience and norms of the interpreter. What if the weight and meaning we attribute to a certain fact—thus helping us successfully interpret on one occasion—proves to be wrong on another? To use the same example as before: Knowing the only dog lurking around makes interpreting the signs very easy. But what if there are other dogs in the neighborhood, with different behavior and habits?

Naturally, the third part of the process, the one of assessment prediction, is probably the most likely to fail. Not only is it based on two previous phases of a process, each prone to flaws and mistakes as detailed above, it can also go wrong due to its own weakness. For even if all the evidence collated is correct and our interpretation thereof is precisely right, there is no guarantee whatsoever that our adversary will act next time the same way he did before.

To sum up, it is quite difficult to give an accurate intelligence assessment. There are at least three steps and in each something can go wrong. These three steps are built upon one other, and a mistake in any one of them can easily disrupt the whole process. What if I misidentified the paw prints? What if they were actually prints of a cat or, on the other hand, a bear? How old are these footprints? And can I be sure in which direction that animal proceeded?

Moreover, in order to make an analysis complete, this triple-step process of evidence accumulation, interpretation, and assessment should refer to two parallel aspects: both capacities and intentions. When analyzing an enemy or opponent one should not forget these are two distinct aspects. To use a recent example from another area, a South American leader may proclaim his intention of bringing an end to U.S. global hegemony, but he has no capacity to do so. The European Central Bank, for its part, has some capacity to do so but seems to have no such intentions at the moment. The Chinese government, on the other hand, might have both the capacity and the intentions. Still, one cannot predict what the Chinese will do.

It might be tempting to model the three parts of the intelligence analysis as parallel to three time phases: collection of evidence from the past, their interpretation in the present, and making an evaluation for the future. While this model might work in some cases, the work of intelligence is usually continuous. New pieces of evidence are streaming in constantly, modifying the picture we see, igniting and fueling the process over and over again. Things become even more complicated when we consider the fact that usually there is a mutual relationship between the researcher and the object researched. Most objects of intelligence analysis are not naive snapshot models, but adversaries aware of being viewed; hence they change their behavior in accordance to our actions. To return to the example I began with, if the dog sees I'm viewing him on his way to the kitchen, he might try to use another way to get there—the back door, for instance. Therefore, in a world where communication and action advance as fast as a television broadcast, evidence collection and decision making create an endless circle, much like a dog chasing its own tail.

And Now to Iraq

In the decade following the 1991 Gulf War, the United Nations was continuously calling upon Iraq for the complete elimination of its WMD. The United

Nations established a special commission of experts (United Nations Special Committee [UNSCOM]), which was active between 1991 and 1999, its mission being to ensure Iraq's compliance with UN resolutions by obliging it to dismantle its WMD. The Iraqi authorities, however, not only failed to meet their disarmament obligations, but impeded UNSCOM's work by preventing weapons inspectors from visiting facilities and by removing documents from them. A serious corroboration to the assumption that Iraq had WMD came in August 1995, as two leading figures in Iraq's weapons industry—the brothers Hussein and Saddam Kamil Hassan al-Magid, cousins and sons-in-law of Saddam Hussein—defected to Jordan. There, they provided UNSCOM with valuable information about Iraq's past WMD projects. Hussein Kamil claimed, however, that he also ordered to destroy all WMD after the 1991 war. In February 1996 the two brothers went back to Baghdad after being told they would be pardoned for their defection, but three days after their return Saddam's security forces came to arrest them. The two brothers were killed in the gun battle that broke out as they resisted their arrest.

Two and a half years later, on October 31, 1998, President Bill Clinton signed the Iraq Liberation Act, supporting Iraqi opposition groups in their efforts to remove Saddam Hussein from power. The following day, the Iraqi government announced it would no longer cooperate with UN weapons inspectors. A year later, in December 1999, UN Secretary General Kofi Annan managed to broker a deal whereby Iraq allowed weapons inspectors back into the country the following year as part of the United Nations Monitoring, Verification and Inspecting Commission (UNMOVIC). While some officials of UNMOVIC were of the opinion that Iraq had dismantled its WMD facilities during the 1990s, in December 2002 UNMOVIC Chairman Hans Blix told the UN Security Council that the Iraqi weapons declaration filed on December 7, 2002, was "essentially a reorganized version" of its reports from 1997, and "not enough to create confidence" that Iraq had abandoned its WMD efforts.

In January 2005, the Iraq Survey Group, a search and research team set up by the Central Intelligence Agency (CIA) and Defense Intelligence Agency (DIA), came to the conclusion that Iraq had quit developing WMD in 1991 and had no WMD at the time of the 2003 invasion (although some remnants of pre-1991 production were found, these were not the weapons for which the United States "went to war"). However, while the search for ready-to-use WMD brought nothing, suspicion of Iraq's innocence still did not dissipate, as Iraq could not offer any credible proof that its WMD program had really been shut down. It might be possible that Iraq's lack of compliance with UN inspections meant they were simply hiding the weapons, either somewhere in Iraq or abroad.

Actually, this is a classical example of the problem of falsifiability: In order to prove the existence of something, one specimen should suffice, but no finite number of failed searches can prove its nonexistence. In our case, finding one

---

UN Weapons Inspector Hans Blix

Hans Blix was the first executive chairman of the United Nations Monitoring, Verification and Inspection Commission, from 2000 until 2003 involved in the search for weapons of mass destruction in Iraq prior to the U.S. invasion.

Born in 1928 in Uppsala, Sweden, his father and his grandfather were both professors, and he completed his undergraduate degree at Uppsala University and then went to Columbia University, before completing his doctorate at the University of Cambridge in Britain and becoming an associate professor in international law. After several postings in the Swedish government, and on his country's delegation to the United Nations, in 1978–1979, he was the Swedish foreign minister.

From 1981 until 1997 Blix was the director general of the International Atomic Energy Agency and was involved in inspections of nuclear facilities around the world, including in Iraq before the reactor Osiraq was destroyed by the Israelis in 1981.

---

barrel of nerve gas in just one Iraqi storage room would have been enough to prove Iraq had WMD; but even 10,000 empty rooms with no trace of chemicals could not offer proof that Iraq had no such substances, because they might be found in the next room.

Conclusion

Although many Americans will find it hard to believe, the United States has become a modern empire, parallel to many other empires in human history. Like all empires, it has an inherent drive to defend its borders and protect its interests in its periphery. And the United States has already done it before, playing a decisive role in the biggest wars the world has ever seen. Though not the only sheriff of the town, the United States is very much one of the world's policemen.

But things have changed in the past half-century. Video killed the radio star, and the New War brought the Total War to an end. Wars today are fought less on the practical battlefield and more in the media and in people's minds and beliefs, especially in wars between two totally unequal forces. Concomitantly, the aims of New Wars are much less definitive than those of Total War. However, as international norms still demand that wars have an aim, warmongers are still looking for them.

In a radio speech on March 22, 2003, four days after the beginning of the invasion of Iraq, President George W. Bush claimed that, for U.S. forces, the "mission is clear, to disarm Iraq of weapons of mass destruction, to end Saddam Hussein's support for terrorism, and to free the Iraqi people." Of these three objectives, two were known to be complete nonsense at the very moment they were stated. First, Saddam Hussein's support of terrorist organizations was minor as compared with that of other state presidents at that time, like Jiang Zemin of China or Bush

himself, for example. Accusations of Saddam's ties with Al Qaeda are dubious, to say the least. Only two weeks before the war, Osama Bin Laden was quoted saying that Saddam was a socialist infidel. Second, there were very few cases in history—to put it mildly—when peoples were freed by the bayonets of foreign armies. Such "good intentions" usually just replace one oppressive occupation with another, frequently the new one being more problematic than the former.

The third reason, however, could not be so easily ruled out, as all three parts of the intelligence analysis showed Iraq might still hold WMD. There was evidence it had bought some materials in the past—obviously it did not purchase them in order to make chocolate pralines—and Saddam's regime has continuously proved to be murderous against civilians and soldiers, both inside Iraq and outside its borders.

Naturally, all this could not prove with total certainty that Iraq was holding WMD, nor that it was going to use them. But here comes the question of risk management. We have serious doubts whether Iraq was holding WMD. But if it was, were we willing to take that risk?

About two centuries before American decision makers faced this decision, it was another U.S. statesman, Benjamin Franklin, who suggested a solution to this problem: "When in doubt, don't." In order to be on the safe side, doubts were ignored in the case of Iraqi WMD. Human beings as we are, we shall always be wiser in retrospect. The question still remains open: What can we learn from the Iraqi case when dealing with other such dilemmas in the not so distant future?

References and Further Reading

Bauman, Zygmunt. *Globalization: The Human Consequences*. Cambridge, UK: Polity Press, 1998.

Brzezinski, Zbigniew. *Second Chance: Three Presidents and the Crisis of American Superpower*. New York: Basic Books, 2007.

Clausewitz, Carl von. *On War*. New York: Penguin, 1982.

Cook, Tim. *No Place to Run: The Canadian Corps and Gas Warfare in the First World War*. Vancouver: University of British Columbia Press, 1999.

Heinberg, Richard. *The Party's Over: Oil, War and the Fate of Industrial Societies*. Gabriola Island, BC: New Society Publishers, 2003.

Hobsbwam, Eric. *The Age of Extremes*. New York: Vintage, 1996.

Kaldor, Mary. *New and Old Wars: Organized Violence in a Global Era*. Stanford, CA: Stanford University Press, 1999.

Maier, Charles. *Among Empires: American Ascendancy and Its Predecessors*. Cambridge, MA: Harvard University Press, 2006.

Popper, Karl. *The Logic of Scientific Discovery*. New York: Basic Books, 1959.

Shultz, George. *Turmoil and Triumph: Diplomacy, Power, and the Victory of the American Ideal*. New York: Scribner, 1995.

Tripp, Charles. *A History of Iraq*. New York: Cambridge University Press, 2002.

United Nations. *Yearbook of the United Nations*. New York: United Nations Department of Public Information, 1986.

CON

The official rationale for the U.S. war against Iraq in March 2003 was the threat of Iraqi possession of weapons of mass destruction (WMD). After the terror attacks in New York and Washington on September 11, 2001, President George W. Bush argued that there was a link between Islamic terrorism and Iraq. He felt that Saddam Hussein was an undeterrable aggressor possessing chemical and biological weapons who would seek any opportunity to kill Americans. In addition, he warned, Iraq was close to acquiring nuclear weapons. In 2002 the Bush administration tried to convince the public of a connection between Al Qaeda leader Osama Bin Laden and Saddam Hussein. It implied that Saddam was cooperating with Al Qaeda and had assisted in the 9/11 attacks. In fact, the mere possibility of such a link sufficed as *casus bellum*. Despite the weakness of the evidence for its claims, the Bush administration succeeded in convincing a majority of the U.S. public that Iraq posed a threat so extreme and immediate that it could be dealt with only by preventive war. "Time is running out on Saddam Hussein," Bush remarked in January 2003. "He must disarm. I'm sick and tired of games and deception" (quoted from Freedman 2004: 31).

However, soon after the U.S. military intervention in March 2003, it became clear that there were no WMD in Iraq. In early 2004, David Kay, the head of a U.S. group surveying Iraq for the evidence of WMD, declared that "we were all wrong" and gave up the search. This section focuses on the Bush administration's rhetoric and the reality of Iraqi WMD in the last few months before the U.S. military intervention in Iraq of March 2003. It is now clear that the Bush team's urge to war came first—and the threat assessments of the U.S. intelligence agencies only later. All available evidence leads to the shocking conclusion that the Bush administration deliberately inflated Iraq's WMD threat (Freedman 2004; Kaufman 2004; Payne 2004). Thus, the U.S.-led war against Iraq, in retrospect, must be classified as an unjustified, aggressive war, in violation of international law.

Iraq and Weapons of Mass Destruction

Iraq started its nuclear weapons program and biological and chemical weapons programs in the 1960s and 1970s. During the Iran-Iraq War in the 1980s, Western

countries maintained good, constructive, and lucrative relations with Iraq—the archenemy of Ayatollah Khomeini's Iran. In 1981, Israel destroyed Iraq's nuclear reactor in Osiraq to prevent a nuclear Iraq. After the 1991 Gulf War, the West combined a strict containment policy of Iraq with coercive elements. A UN mission was tasked with implementing Iraq's disarmament demands of UN Security Council Resolution 687 of April 1991. In 1998 the UN disarmament process collapsed because of continuing Iraqi obstruction (Freedman 2004: 9–14).

During the U.S. presidential election year of 2000, the republicans called for a "comprehensive plan for the removal of Saddam Hussein," but it was not a big campaign issue. George W. Bush prophetically promised that if Saddam was caught "developing WMD in any way, shape, or form, I'll deal with that in a way that he won't like" (quoted from Lemann 2001). Yet, his main foreign policy adviser at that stage, Condoleezza Rice, was more relaxed and said that in her view, Saddam would be deterrable even with WMD (Rice 2000). Prior to 9/11, Iraq was not singled out as a direct threat to U.S. national interests. In March 2001, U.S. Secretary of State Colin Powell described international sanctions against Iraq as a success. He added: "Even though we know Saddam Hussein is working on weapons of mass destruction, we know he has things squirreled away, at the same time we have not seen that capacity emerge to present a full-fledged threat to us" (Powell 2001).

From 1991 to 2002, the U.S. intelligence community did not believe Iraq was a serious threat to the United States. Even as late as December 2001, a National Intelligence Estimate (NIE) declared that "Iraq did not appear to have reconstituted its nuclear weapons program" (Senate Intelligence Report 2004: 85). The Iraqi threat was only inflated in 2002, in order to justify a regime change in Iraq and a U.S. military invasion.

Saddam the Terrorist: Inventing Links to 9/11 and Al Qaeda

The terrorist attacks on 9/11 radically changed U.S. policy toward Iraq. President Bush realized that the world had changed and that the containment of Iraq was no longer an option. According to journalist Bob Woodward, Bush said on September 17, 2001, "I believe Iraq was involved in the 9/11 attacks, but I'm not going to strike them now." He added: "I don't have evidence at this point" (Corn 2003). The Bush government retaliated against the Taliban in Afghanistan in October 2001, but the main focus of the "war on terror" soon moved to Iraq. Secretary of Defense Donald Rumsfeld had requested plans for striking Iraq barely five hours after the Pentagon had been hit on 9/11 (Freedman 2004: 34).

The assertion that Hussein was personally responsible for assisting in the 9/11 attacks was based on an alleged meeting between Mohammed Atta, the pilot of one of the planes that hit the Twin Towers, and an Iraqi intelligence officer in

Prague in April 2001. The source for this was a single Czech informant whom Czech intelligence reported was not credible (Isikoff 2002; Tagliabue 2001). The Federal Bureau of Investigation (FBI) and the Central Intelligence Agency (CIA) also quickly concluded that no such meeting had taken place. Evidence suggested that Atta had been in Virginia at the time in question (Simpson 2003; Tyler 2002). Yet, the Bush administration continued to repeat the story despite the lack of evidence. Vice President Dick Cheney, for example, stated in September 2002, "We have reporting that places Atta in Prague with a senior Iraqi intelligence official a few months before September 11. It is credible" (Cheney 2002).

The Bush administration's second main claim linking Iraq to Al Qaeda concerned Abu Musab al-Zarqawi, the head of a terrorist group called al-Tawhid. Rumsfeld said that he had "bulletproof evidence" of a link between Hussein and Al Qaeda (Schmitt 2002). The fullest version of this claim came in Powell's famous speech to the UN Security Council on February 5, 2003, where he asserted that Iraq was "harboring" Zarqawi, "a collaborator of Osama bin Laden" and had allowed him to establish a base of operations in Baghdad for Al Qaeda affiliates, that the Zarqawi network helped establish a poison and explosive training center in a camp belonging to an Islamist group called Ansar al-Islam (Powell 2003).

Throughout 2002, the Bush team placed the looming war against Iraq firmly under the heading of the war on terrorism. In January 2002, President Bush claimed that Iraq, Iran, and North Korea were part of an "axis of evil" that posed a "grave and growing danger" because of their pursuit of WMD (Bush 2002a). Bush and his key national security advisers over and over again mentioned Saddam Hussein, Al Qaeda, and Iraq together in the same sentence. Bush argued in a speech in Cincinnati on October 7, 2002: "The danger is that they work in concert. The danger is, that the Qaeda becomes an extension of Saddam's madness and his hatred and his capacity to extend weapons of mass destruction around the world. Both of them need to be dealt with . . . you can't distinguish between the Qaeda and Saddam when you talk about the war on terror" (Bush 2002d). On November 7, 2002, Bush reemphasized: "Saddam Hussein is a threat because he is dealing with the Qaeda" (Bush 2002f). In his 2003 State of the Union address, Bush said, "Imagine those 19 hijackers with other weapons and other plans—this time armed by Saddam Hussein" (Bush 2003a).

The U.S. president kept repeating this message until a few days before the launching of the Iraq War. "Iraq has aided, trained and harbored terrorist, including operatives of the Qaeda," Bush said on March 17, 2003. In strong words, he asserted: "The danger is clear: using chemical, biological, or, one day, nuclear weapons provided by Iraq, the terrorists could one day kill hundreds of thousands of people in our country or any other" (Bush 2003c).

In sum, the Bush administration encouraged public belief that Saddam Hussein had been involved in the 9/11 atrocities. And it was successful in swaying U.S. public opinion. "Polls in late 2002 showed that 70–90 percent of the

American public believed that [Saddam] Hussein would sooner or later attack the United States with weapons of mass destruction. Between 45 and 66 percent also believed that he had assisted the [9/11 attackers]" (Kaufman 2004: 30). Interestingly, however, right after 9/11, only 3 percent had thought Iraq or Saddam Hussein was behind the terrorist attacks (Freedman 2004: 20).

Inflating the Threat: Iraq's Alleged WMD Program

Serious public discussion about Iraq began only in late summer 2002. Vice President Dick Cheney launched the debate in a widely noted speech to the Veterans of Foreign Wars on August 26, 2002—several weeks before the intelligence community would produce its now famous estimate on Iraq. Pointing at information obtained from Iraqi defectors, Cheney declared with great certainty that "simply stated, there is no doubt that Saddam Hussein now has weapons of mass destruction. There is no doubt he is amassing them to use against our friends, against our allies, and against us. We now know that Saddam has resumed his efforts to acquire nuclear weapons. Many of us are convinced that Saddam will acquire nuclear weapons fairly soon" (Cheney 2002).

Bush's National Security Adviser Condoleezza Rice chimed in with a very pessimistic estimate on the Iraqi threat. "The problem here is that there will always be some uncertainty about how quickly Saddam Hussein can acquire nuclear weapons," Rice said in an interview with CNN. "But we don't want the smoking gun to be a mushroom cloud" (Rice 2002). And in a well-timed address at the UN General Assembly on September 12, 2002—with America still mourning the first anniversary of the 9/11 terror attacks—Bush declared, "The first time we may be completely certain Saddam Hussein has nuclear weapons is when, God forbid, he uses one" (Bush 2002b).

Secretary of Defense Rumsfeld also participated in the competition among Bush officials to find spectacular metaphors for convincing public opinion of the certainty of the Iraqi threat. "There's no debate in the world as to whether Iraq has those weapons. There's no debate in the world as to whether Iraq continues to develop and acquire them. We all know that. A trained ape knows that. All you have to do is read the newspaper," Rumsfeld said on September 13, 2002 (Rumsfeld 2002).

In its public campaign, the Bush administration officials did not talk about abstract risks but concrete threats, giving specific time references. In mid-September, President Bush declared: "Should Saddam Hussein's regime acquire fissile material, it would be able to build a nuclear weapon within a year" (Bush 2002c). And Vice President Cheney argued just days before the war was launched that Hussein was "absolutely devoted to trying to acquire nuclear weapons. And we believe he has, in fact, reconstituted nuclear weapons" (Cheney 2003).

A key document in the run-up to the Iraq War was the national intelligence estimate dated October 1, 2002, which was the most important summary prepared

UN weapons inspectors carry equipment to their jeep prior to leaving Baghdad UN headquarters March 6, 1998. The men were part of a group of UN weapons inspection teams that conducted spot tours of suspected weapons sites. (AP Photo/Peter Dejong)

by the intelligence community. It argued that Iraq was "reconstituting its nuclear program," had chemical and biological weapons, and that all key aspects of Iraq's offensive biological weapons program were active and most elements were larger and more advanced than they had been before the 1991 Gulf War. Yet, a 2002 national intelligence estimate in fact had concluded that Saddam Hussein was unlikely to initiate an unprovoked WMD attack against the United States. Thus, they did not share the overly pessimistic rhetoric that the Bush team had applied in speeches around the first anniversary of the 9/11 terror attacks.

In 2002 the Bush administration was able to achieve unanimous international support for UN Security Council Resolution 1441. However, Bush and his junior partner Tony Blair, the British prime minister, failed to win a second UN Security Council resolution specifically authorizing war against Iraq. France, Germany, and Russia preferred continuing weapons inspections and dissented to the United States's rush to war. They did not buy the alleged link of Saddam Hussein with 9/11 (Dombrowski and Payne 2003).

From October 2002, the central argument for fighting a war in Iraq was disarmament. Powell emphasized: "All we are interested in is getting rid of these weapons of mass destruction" (Powell 2002). In October 2002, Bush stated (using the famous phrase that Rice had used a few weeks earlier): "Facing clear evidence of peril, we cannot wait for the final proof—the smoking gun—that

could come in the form of a mushroom cloud. We cannot stand by and do nothing while dangers gather" (Bush 2002e). The various statements made by the Bush administration from August 2002 until mid-March 2003 created the very strong—and ultimately false—impression that Iraq had an active and dangerous nuclear weapons program that was precariously close to success.

The only significant concrete claims made by the Bush administration with regard to Iraqi attempts to reconstitute its nuclear weapons program was discredited by internal analyses before it was made public. Rice, Cheney, Bush, and Powell all "asserted that Iraq was trying to import high-quality aluminum tubes that were only really suited for nuclear weapons programs, centrifuge programs. CIA, Department of Energy, and State Department analysis, however, had concluded that the evidence did not support that interpretation." In addition, the "IAEA [International Atomic Energy Agency]pointed out that Iraq's centrifuge program had abandoned the use of the alloy in question before 1991; that the tubes had coatings suitable for use in rockets but not for centrifuges and that the number ordered—120,000—was wildly excessive for uranium enrichment but consistent with artillery use" (Kaufman 2004: 25).

Perhaps the best-known distortion of the intelligence related to alleged Iraqi import of tons of uranium was from Niger. "The British government learned that Saddam Hussein recently sought significant quantities of uranium from Africa," Bush declared in his 2003 State of the Union address (Bush 2003a). It soon turned out that documents had been forged and that the CIA had previously convinced the White House to remove a similar assertion from the president's October 2002 speech in Cincinnati. The IAEA also determined that the documents were not authentic (Pincus 2003).

David Kelly and Iraqi Weapons of Mass Destruction

During a parliamentary committee hearing into the British government producing its dossier on weapons of mass destruction (nicknamed the 'dodgy dossier'), it emerged that some comments made by the journalist Andrew Gilligan came from information from David Kelly (1944–2003).

Kelly was from a Welsh mining family and had studied at the University of Leeds and then the University of Birmingham, before completing a doctorate in microbiology at the University of Oxford. He had then become an adviser to the Ministry of Defence, and worked on biological warfare at Porton Down. At the end of the Cold War he was involved in discovering supplies of biological weapons in the former Soviet Union, and this led to his appointment to United Nations Special Committee (UNSCOM).

Visiting Iraq many times, David Kelly often used his spare time in Baghdad perusing book shops. He also became a Baha'i. After appearing in the spotlight at the parliamentary hearing, he returned home and went for a walk one morning and later was found dead. The subsequent inquest showed that he had committed suicide.

In March 2003, Bush said that Iraq was moving biological and chemical weapons "every 12 to 24 hours" (Bush 2003b). Similarly, Powell presented a set of photographs to the UN Security Council in February 2003 that purportedly showed mobile biological weapons laboratories in trucks (Powell 2003). The "bioweapons laboratories" later turned out to be gas generators for filling weather balloons (Kaufman 2004: 29).

"Before the war, four months of unrestricted, essentially unhindered IAEA inspections from December 2002 to March 2003 eliminated virtually all remaining doubt" (Kaufman 2004: 25). After three months of intrusive inspections, IAEA director Mohammed El Baradei concluded that there was "no evidence or plausible indication of the revival of a nuclear weapons program in Iraq" (El-Baradei 2003).

By February 2003, the highly reputed *Washington Post* had completely accepted the perception that Iraq had WMD and echoed the Bush administration's call for war in an editorial titled "Irrefutable," printed the day after Powell's UN presentation. The *Post* wrote: "It is hard to imagine how anyone could doubt that Iraq possesses weapons of mass destruction." On the same day, *Washington Post* journalist Mary McGrory confessed about Powell's speech: "He persuaded me" (quoted from Solomon 2005: 46). The Bush administration had also succeeded in convincing a majority of Americans of its dubious justification for war against Iraq. "Prewar polls showed that 55 to 69 percent of Americans believed that Saddam Hussein already possessed WMD, and more than 95 percent believed that he was building them. In one poll, 69 percent believed that Iraq already had nuclear weapons, and in another, 80 percent thought this likely" (Kaufman 2004: 30).

On March 20, 2003, the Bush administration launched its war against Iraq. A vast majority of Americans saw the war as retaliation against the 9/11 attacks and a just war to destroy the alleged Iraqi nuclear weapons program. The official reason for war was not regime change, it was not democratizing Iraq, or getting rid of a horrible tyrant. The major issue was nuclear disarmament, as President Bush made clear in his last prewar press conference: "Our mission is clear in Iraq. Should we have to go in, our mission is very clear: disarmament" (Bush 2003b).

After the Invasion I: What about Saddam's Links with Bin Laden?

Relatively soon after the U.S. invasion of Iraq, the two main reasons for waging war against Saddam Hussein—WMD and links to Al Qaeda—collapsed like a cardboard castle. The propaganda strategy of the Bush administration backfired when their bluff was exposed in Iraq.

> Zarqawi and Ansar al-Islam may indeed have been allied to the Qaeda; since the US invasion of Iraq, Zarqawi has sought the Qaeda's help in organizing

resistance to the occupation. There was no evidence, however, that Ansar or Zarqawi had been cooperating with Hussein. On the contrary, Ansar was formed in 2001 for the purpose of overthrowing Hussein's regime and transforming Iraq into an Islamic state. It operated in an area of northern Iraq not under Baghdad's control. (Kaufman 2004: 18)

U.S. intelligence officials "had complained both privately and publicly that the evidence did not support administration claims" (Smith 2003). Powell retracted the Zarqawi claims in 2004. In contrast to the certainty asserted earlier, he now stated: "I have not seen smoking-gun, concrete evidence about the connection between Iraq and the Qaeda. But I think the possibility of such connections did exist, and it was prudent to consider them" (Marquis 2004).

Similarly, President Bush also had to admit that Iraq had no ties to the 9/11 attacks. Bush confessed in September 2003, "We've had no evidence that Saddam Hussein was involved with September the 11th" (Bush 2003d). But he maintained in 2004, that Hussein "had the capacity to make a weapon and let that weapon fall into the hands of a shadowy terrorist network" (Bush 2004b).

The official 9/11 Commission also reported in mid-2004 "to date we have seen no evidence that contacts between Iraqi officials and the Qaeda members ever developed into a collaborative operational relationship. Nor have we seen any evidence that Iraq cooperated with the Qaeda in developing or carrying out any attack against the United States" (9/11 Commission Report 2004: 66).

The idea of significant cooperation between Saddam Hussein and the Qaeda had always been "implausible, because the Iraqi Ba'athists were exactly the sort of secularists that Osama bin Laden blamed for the corruption of Muslim society and for its defeats at the hands of Christians and Jews." Bin Laden had consistently denounced Hussein in the strongest terms (Kaufman 2004: 19). Hussein, on the other hand, had limited his support for international terrorism to "payments to families of Palestinian suicide bombers, a practice also common in other Arab states" (Kaufman 2004: 19). Ironically, Syria, Iran, Yemen, Saudi Arabia, and Pakistan provided far more practical support to terrorism than did Hussein's Iraq (Risen 2002).

Thus, there was no basis at all for the allegations made by the Bush administration that Saddam Hussein would give weapons of mass destruction to terrorists who might use them against the United States. Still, public belief in the administration's prewar statements declined only slightly. In polls taken in April 2004, 47 percent still believed that "clear evidence that Iraq was supporting the Qaeda had been found in Iraq," while 36 percent thought that Hussein was personally involved in the 9/11 attacks (Kaufman 2004: 31-2.).

After the Invasion II: Where Have All the WMD Gone?

In March 2003, the United States sent special teams into Iraq to find Saddam Hussein's weapons of mass destruction—the 75th Exploitation Task Force and

the covert Task Force 20. In the first weeks of the war, high-ranking officials still pretended that the United States would find the WMD. Secretary of Defense Rumsfeld even claimed, "They're in the area around Tikrit and Baghdad and east, west, south, and north somewhat" (Rumsfeld 2003). That statement was a plain lie. In fact, despite the administration's repeated assertions about the great and growing threats from Iraqi arsenals, it soon became clear that the world's most dangerous weapons were not present in Iraq.

The failure of U.S. search teams to find evidence of a reconstituted Iraqi nuclear weapons program since 1991 have validated the IAEA prewar findings. The U.S. chief weapons inspector David Kay left Iraq in December 2003 and resigned from his post a month later. Apparently, he considered the search for alleged WMD to be a pointless waste of time. In early 2004, he gave countless interviews and speeches stating that Iraq had no weapons to find and that an enormous prewar intelligence failure had concurred. "We were all wrong," Kay confessed (quoted from Payne 2004:14–5). In July 2004, Kay declared that the Bush administration should abandon its "delusional hope" of finding such weapons in Iraq (Zakaria 2004).

His successor, Charles Duelfer, presented the final report of the Iraq Study Group in September 2004. It stated that while Saddam Hussein badly wanted various unconventional weapons, Iraq had destroyed its chemical stockpile and ended its nuclear program in 1991. In addition, the report says, Iraq had abandoned its biological program in 1996 after the destruction of its key research facility. The Duelfer report concluded that Hussein wanted such weapons to ensure Iraq's position in the Middle East and to threaten Iran—not to attack the United States (CIA 2004).

The Senate Select Committee of Intelligence concluded "the major key judgments in the NIE . . . either overstated, or were not supported by, the underlying intelligence provided to the Committee." Also, the committee felt that the NIE did not adequately explain the uncertainties surrounding the quality of the evidence. And it found that the conclusions of the intelligence community had reflected pessimistic biases and serious mistakes of tradecraft (Senate Select Committee 2004). "Analysts from the Carnegie Endowment for International Peace and other nongovernmental organizations that have reviewed the publicly available data also concluded that the case for war against Iraq was built on very weak evidence" (Payne 2004: 10).

In early 2004, President Bush had to admit in an address before the joint session of the Congress that hundreds of highly trained U.S. inspectors, the Iraq Survey Group, had only been able to assemble evidence of "dozens of weapons of mass destruction–related program activities" rather than actual Iraqi weapons or weapons-production infrastructure (Bush 2004a). More honestly, Colin Powell would later describe it in his speech on Iraq's possession of WMD to the UN Security Council as a "blot" on his professional career. When asked about the

speech during an interview with Barbara Walters in September 2005, Powell said: "The speech will always be a part of my record. It was painful. It's painful now" (Powell 2005).

The Bush administration's claims that Iraq possessed large stocks of chemical and biological weapons and was continuing to produce them were also inflated. U.S. inspections after the 1991 showed that it was unlikely that Iraq had chemical or biological weapons or active weapons programs. In 2004, David Kay stated, "I don't think they existed" (quoted from Zakaria 2004).

Thus, none of Iraq's WMD programs operated at more than a planning level after the 1991 Gulf War. Again, the U.S. public was slow in accepting the fact that the Bush government had deliberately lied to them. In April 2004, polls indicated that 57 percent of Americans still believed that Iraq had possessed WMD, while 52 percent thought that Iraq still had WMD that had not yet been found (Kaufman 2004: 31).

Conclusion

The primary rationale for the U.S. attack on Iraq was Saddam Hussein's alleged nuclear, chemical, and biological weapons stockpiles and programs. Iraq's suspected connections to international terrorism also played an important role in justifying the U.S. action. The Bush administration had spent much of 2002 emphasizing the key role of Iraq in their "war on terror" to justify their plan to topple Saddam Hussein's government by military force. It deliberately—but erroneously—linked Saddam's Iraq with the terrorist attacks against the United States of 2001, knowing very well that a traumatized U.S. public would support retaliation against the terrorists responsible for the 9/11 atrocities. It portrayed Saddam Hussein as an evil madman bent on the destruction of the United States. The deception and the public lies of the Bush neoconservatives worked. In March 2003, a majority of Americans believed Saddam was responsible for 9/11 and thus supported the Iraq War.

The Bush government also inflated the threat resulting from Iraq's alleged WMD stockpiles and programs. According to Paul Wolfowitz, then deputy secretary of defense, the Bush administration "settled on the one issue that everyone could agree on which was weapons of mass destruction as the core reason for war" (Wolfowitz 2003). In retrospect, it seems clear that the Bush government greatly distorted the prewar debate about Iraq and overstated some of the intelligence community's key assessments of the Iraqi threat. The White House employed dubious rhetorical strategies that overemphasized alleged WMD in the public debate.

Despite the numerous public assertions by the Bush administration about great and growing threats from Iraqi arsenals, there were no WMD in Iraq. The case for war against Iraq was built on very weak evidence. President Bush was

certainly aware of this fact. In December 2002, when he received a full briefing on the quality of the intelligence on Iraq, Bush was unimpressed. "It's not something that Joe Public would understand," he said and asked CIA Director George Tenet if this was "the best we've got." Reportedly, Bush was told: "It's a slam dunk case" (Woodward 2004: 247–50).

By greatly exaggerating the status of the Iraqi WMD program, the Bush administration deliberately misled the general public and mass media. Why did the American public and media uncritically buy the "war propaganda" of the Bush government? The shock of the 9/11 atrocities created a crisis atmosphere that reduced public skepticism about both diagnoses of threats and proposed solutions. The Bush administration played with these post-9/11 fears. U.S. Secretary of State Colin Powell, for example, explained in his UN speech on February 5, 2003: "Leaving Saddam Hussein in possession of weapons of mass destruction for a few more months or years is not an option, not in a post-September 11th word" (Powell 2003).

Still, the Iraq War was a "war by choice," not a "war by necessity." When the bombs began falling on Iraq on March 20, 2003, the Bush administration aimed at regime change and implementing their (unrealistic) vision of a democratic Greater Middle East. Disarmament and WMD was not in the forefront of U.S. thinking. The IAEA had made clear after their onsite inspections from January to March 2003 that Iraq did not pose an immediate threat. But it would take a few years and the loss of more than 4,000 U.S. soldiers (not to mention the lost lives of soldiers who allied with the United States) in the Iraq War, until the majority of Americans began to realize that the Bush government had manipulated the U.S. public by playing the 9/11 card to sell the Iraq War. And some administration officials continued to use the prewar claims after the invasion and against all evidence. Vice President Dick Cheney, for example, still declared in April 2004 that there was "conclusive evidence that Hussein did have programs for weapons of mass destruction" as well as "overwhelming evidence that there was a connection between the Qaeda and the Iraqi government" (Cheney 2004).

References and Further Reading

9/11 Commission Report. 2004. Available at www.gpoaccess.gov/911/pdf/fullreport.pdf (accessed June 24, 2010).

Bush, George W. All quotes from http://www.presidency.ucsb.edu:

Bush, George W., 2002a. State of the Union Address. January 29.

Bush, George W., 2002b. Remarks at the UN General Assembly. September 12.

Bush, George W., 2002c. Radio Address. September 14.

Bush, George W., 2002d. Remarks. September 27.

Bush, George W., 2002e. Address to the Nation on Iraq. Cincinnati, Ohio. October 7.

Bush, George W., 2002f. News Conference. November 7.

Bush, George W., 2003a. State of the Union Address. January 28.

Bush, George W., 2003b. Press Conference. March 6.

Bush, George W., 2003c. Address to the Nation on War with Iraq. March 17.

Bush, George W., 2003d. Remarks. September 17.

Bush, George W., 2004a. Address on the State of the Union. January 20.

Bush, George W. NBC *Meet the Press*. February 8, 2004b. Available at http://www.msnbc.msn.com/id/4179618/ (accessed August 19, 2010).

Central Intelligence Agency. *Comprehensive Report of the Special Advisor to the DCI on Iraq's WMD*. September 30, 2004. Available at https://www.cia.gov/library/reports/general-reports-1/iraq_wmd_2004/index.html (accessed August 19, 2010).

Cheney, Dick. Address at VFW 103rd National Convention. August 26, 2002. Available at http://georgewbush-whitehouse.archives.gov/news/releases/2002/08/20020826.html (accessed August 19, 2010).

Cheney, Dick. NBC *Meet the Press*. March 16, 2003. Available at http://www.msnbc.msn.com/id/3080244/ (accessed August 19, 2010).

Cheney, Dick. NPR *Morning Edition*. January 22, 2004. Available at http://www.npr.org/templates/story/story.php?storyId=1610113 (accessed August 19, 2010).

Cirincione, Joseph, Jessica T. Methews, and George Perkovich, with Alexis Orton. *WMD in Iraq: Evidence and Implications*. Washington, DC: Carnegie Endowment for International Peace, 2004.

Cochran, John. "Reason for War? White House Officials Say Privately the September 11 Attacks Changed Everything." April 25, 2003. Available at http://abcnews.go.com/Nightline/story?id=128467&page=1 (accessed June 24, 2010).

Corn, David. "Who's In Charge?" *The Nation* (February 13, 2003). Available online at http://www.thenation.com/article/whos-charge?page=0,1&comment_sort=ASC (accessed August 19, 2010).

David, Ian. *Unravelling the Know Unknowns: Why No Weapons of Mass Destruction Have Been Found in Iraq*. British American Security Information Council, Special Report, January 2004. Available at http:www.basicint.org/pubs/Research/2004WMD.pdf (accessed June 24, 2010).

Dombrowski Peter J., and Rodger A. Payne. "Global Debate and the Limits of the Bush Doctrine." *International Studies Perspectives* 4 (2003): 395–408.

El-Baradei, Mohammed. "Status of Nuclear Weapons in Iraq." March 7, 2003. Available at www.un.org/News/dh/iraq/elbaradei-7mar03.pdf (accessed June 24, 2010).

Freedman, Lawrence. "War in Iraq: Selling the Threat." *Survival* 46, 2 (2004): 7–50.

Isikoff, Michael. "The Phantom Link to Iraq: A Spy Story Tying Saddam to 9-11 Is Looking Very Flimsy." *Newsweek* (April 28, 2002), web exclusive.

Kaufman, Chaim (2004). "Threat Inflation and the Failure of the Marketplace of Ideas: The Selling of the Iraq War." *International Security* 29 (2004): 5–48.

Kay, David. Testimony before the Senate Armed Services Committee. January 28, 2004. Iraqi Weapons of Mass Destruction Programs. Available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_senate_hearings&docid=f:96675.pdf (accessed August 19, 2002).

Lemann, Nicholas. "The Iraq Factor: Will the New Bush Team's Old Memories Shape Its Foreign Policy?" *New Yorker* (January 22, 2001).

Marquis, Christopher. "Powell Admits No Hard Proof in Linking Iraq to al Qaeda." *New York Times* (January 9, 2004).

Payne, Rodger A. "Deliberating Preventive War." Paper presented at Carnegie Endowment for the International Peace, February 7–8, 2004.

Pincus, Walter. "White House Backs Off Claim on Iraqi Buy." *Washington Post* (July 8, 2003).

Powell, Colin. "Colin Powell on Iraq, Race, and Hurricane Relief." 20/20 ABC News. September 8, 2005. Available at http://abcnews.go.com/2020/Politics/story?id=1105979&page=1 (accessed August 19, 2010).

Powell, Colin. NBC's *Meet the Press*. October 20, 2002.

Powell, Colin. Remarks at the UN Security Council. February 5, 2003. Available at http://www.un.org/News/Press/docs/2003/sc7658.doc.htm (accessed August 19, 2010).

Powell, Colin. Testimony before the Senate Foreign Relations Committee. March 8, 2001. Available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_senate_hearings&docid=f:72227.pdf (accessed August 19, 2010).

Rice, Condoleezza. "Campaign 2000: Promoting the National Interest." *Foreign Affairs* 79:1 (2000).

Rice, Condoleezza. CNN *Late Edition with Wolf Blitzer*. September 8, 2002. Available at www.cnn.com/transcripts/0209/08/le.00.html (accessed June 24, 2010).

Risen, James. "Terror Acts by Baghdad Have Waned, U.S. Aides Say." *New York Times* (February 6, 2002).

Rumsfeld, Donald. Media Roundtable with BBC and VOA. September 13, 2002. Available at www.defenselink.mil/transcripts/transcript.aspx?transcriptid=3652 (accessed June 24, 2010).

Rumsfeld, Donald. Remarks on ABC's *This Week with George Stephanopoulos*. March 30, 2003. www.defenselink.mil/transcripts/transcript.aspx?transcriptid =2185 (accessed June 24, 2010).

Schmitt, Eric. "Rumsfeld Says U.S. Has 'Bulletproof' Evidence of Iraq's Links to al Qaeda." *New York Times* (September 28, 2002).

Senate Intelligence Report. "Foreign Missile Developments and the Ballistic Missile Threat Through 2015." Available at http://www.dni.gov/nic/PDF _GIF_otherprod/missilethreat2001.pdf (accessed August 19, 2010).

Senate Select Committee on Intelligence. *Report on the U.S. Intelligence Community's Prewar Intelligence Assessments on Iraq.* Washington, DC: Government Printing Office, 2004.

Simpson, Can. "Al Qaeda Allegation Is Weakest Link in Justifying War." *Chicago Tribune* (March 20, 2003).

Smith, Michael. "Spies Force Retreat on 'al Qaeda Link.'" *Telegraph* (February 4, 2003).

Solomon, Norman. *War Made Easy: How Presidents and Pundits Keep Spinning Us to Death.* Hoboken, NJ: Wiley, 2005.

Tagliabue, John. "No Evidence Suspect Met Iraqi in Prague." *New York Times* (October 20, 2001).

Tyler, Patrick E. "U.S. Drops Last Link of Iraq to 9/11." *New York Times* (May 2, 2002).

Wolfowitz, Paul. Interview with Sam Tannenhaus. *Vanity Fair* (May 9, 2003). Available at www.defenselink.mil/transcripts/transcript.aspx?transcriptid=2594 (accessed June 24, 2010).

Woodward, Bob. *Plan of Attack.* New York: Simon and Schuster, 2004.

Woodward, Bob, and Dan Balz. "Combating Terrorism: 'It Starts Today.'" *Washington Post* (February 1, 2002).

Zakaria, Tabassum. "Give Up 'Delusional Hope' of Iraq WMD, Kay Says." July 28, 2004. Available at http://www.fantasticforum.com/1res/printthread.php?s =98ef1b7f3f8d9a8b0b0d907be7048cbd&threadid=8625 (accessed August 19, 2010).

Index

About the Editor

STEVEN L. DANVER is a full-time faculty member in the Center for Under-graduate Studies at Walden University. He is a historian of the American West, American Indian peoples, and the interaction of people with the environment. He earned his bachelor's in religious studies at the University of California, Santa Barbara; his master's in historical studies from Graduate Theological Union in Berkeley, California; and his doctorate in history at the University of Utah. His dissertation, *Liquid Assets: A History of Tribal Water Rights Strategies in the American Southwest*, to be published by the University of Oklahoma Press, examines the long history of one of the most important issues of modern relevance to American Indians in the West.

Among the many historical reference projects he has edited or co-edited are ABC-CLIO's *The Great Depression and New Deal: A Thematic Encyclopedia, Seas and Waterways of the World: An Encyclopedia of History, Uses, and Issues,* and a four-book set on *Popular Controversies in World History*. He is currently completing work on an encyclopedia on *Revolts, Protests, Demonstrations, and Rebellions in American History*. Since 2004 he has been managing editor of *Journal of the West*, a respected journal on the history of the North American West for nearly fifty years, and has always sought to publish articles that are both academically sound while at the same time interesting to the non-academic readership.