

A faint, circular portrait of David Hume is visible in the background of the cover. He is shown from the chest up, wearing a dark coat and a white cravat, with his hair styled in a wig.

HUME'S

EPISTEMOLOGY & METAPHYSICS

an introduction

Georges Dicker

**Also available as a printed book
see title verso for ISBN details**

HUME'S EPISTEMOLOGY AND METAPHYSICS

"This is an excellent book. Dicker concentrates squarely on Hume's central arguments, usefully assessing them as contributions to contemporary philosophical debates. He also pays scrupulous attention to the texts, taking account of recent Hume scholarship but remaining attentive to the needs of inexperienced students."

Professor Gary Iseminger
Carleton College

Many of the current philosophical problems over meaning, knowledge, causality, and sense perception can be traced to David Hume's writings in his *Treatise of Human Nature* and *Enquiry Concerning Human Understanding*.

Hume's Epistemology and Metaphysics: An Introduction provides a clear, concise, and accessible guide to the key themes in Hume's philosophy. Issues discussed include Hume's argument that there can be no purely rational demonstration of anything's existence, so that God cannot be proven to exist; that all our scientific knowledge rests on inferences from past experiences that cannot be rationally justified; and that we cannot talk of causality apart from natural science. Georges Dicker reveals the contemporary significance of these problems by clearly and sharply analyzing Hume's reasoning. Throughout, Hume's arguments are also placed against a historical background providing us with essential insight into his criticisms of rationalism and his central place as a founder of empiricism.

Key features of the book also include discussion of Kant's responses to Hume and consideration of more recent responses to Hume's philosophy, allowing the full significance of his thought for contemporary philosophy to emerge.

Accessible to anyone coming to Hume's philosophy for the first time, *Hume's Epistemology and Metaphysics: An Introduction* provides an ideal guide to the main themes in his writing.

Georges Dicker is Professor of Philosophy and Chair of the Philosophy Department at the State University of New York College at Brockport. He is author of *Descartes: An Analytical and Historical Introduction* (1993), *Perceptual Knowledge: An Analytical and Historical Study* (1980), and *Dewey's Theory of Knowing* (1976).

HUME'S
EPISTEMOLOGY AND
METAPHYSICS

An Introduction

Georges Dicker



London and New York

First published 1998
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

This edition published in the Taylor & Francis e-Library, 2001.

© 1998 Georges Dicker

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data

Dicker, Georges, 1942–

Hume's epistemology and metaphysics: an introduction/Georges Dicker.

p. cm.

Includes bibliographical references and index.

ISBN 0-415-16318-8 (hb: alk. paper).

1. Hume, David, 1711–1776 – Contributions in theory of knowledge.
2. Hume, David, 1711–1776 – Contributions in metaphysics.
3. Knowledge, Theory of – History – 18th century.
4. Metaphysics – History – 18th century.

I. Title.

B1499.K7D53 1998

192–dc21 97–34891

CIP

ISBN 0-415-16318-8 (hbk)
ISBN 0-415-16319-6 (pbk)
ISBN 0-203-02071-5 Master e-book ISBN
ISBN 0-203-17461-5 (Glassbook Format)

To Marjorie

CONTENTS

<i>Preface</i>	ix
<i>Acknowledgments</i>	xi
<i>Note on References and Abbreviations</i>	xii
Introduction	1
1 Hume's theory of meaning and its implications	5
1 <i>Hume's theory of impressions and ideas and meaning-empiricism</i>	5
2 <i>Application of Hume's test for meaning to "substance" and "self"</i>	15
3 <i>An alternative to the substance theory</i>	21
4 <i>Critique of Hume's bundle theory of the self</i>	31
2 Hume's theory of knowledge (I): "Hume's Fork"	35
1 <i>Introduction</i>	35
2 <i>"Relations of ideas" and "matters of fact"</i>	35
3 <i>A modernized version of Hume's Fork</i>	41
4 <i>A critique of the modernized "Fork"</i>	49
5 <i>Hume's denials concerning matters of fact</i>	55
3 Hume's theory of knowledge (II): causal reasoning and the problem of induction	61
1 <i>Introduction</i>	61
2 <i>Hume's critique of causal reasoning</i>	62
3 <i>The problem of induction</i>	73
4 <i>Hume's psychological explanation of induction</i>	89
5 <i>Need induction be justified?</i>	91

CONTENTS

4 Hume's theory of causality	99
1 <i>Introduction</i>	99
2 <i>Necessary connection</i>	100
3 <i>Hume's analysis of causation</i>	110
4 <i>The regularity theory: some objections and replies</i>	116
5 Hume's critique of the causal principle	133
1 <i>Introduction</i>	133
2 <i>Why the causal principle is neither self-evident nor demonstrable</i>	134
3 <i>Stroud's critique of Hume's argument</i>	140
4 <i>Kant's "answer" to Hume</i>	143
6 The belief in the existence of body	154
1 <i>Introduction</i>	154
2 <i>Three assumptions behind Hume's account</i>	154
3 <i>The general nature of Hume's account</i>	161
4 <i>The nature and origin of the belief in the existence of body</i>	163
5 <i>The "vulgar system" and the "philosophical system"</i>	174
6 <i>A criticism of Hume's account</i>	178
7 <i>A contemporary perspective on Hume's account</i>	183
8 <i>A Kantian response to scepticism regarding the senses</i>	190
<i>Notes</i>	195
<i>Bibliography</i>	207
<i>Index</i>	211

PREFACE

The purpose of this book is to present and assess David Hume's most influential contributions to epistemology and metaphysics in a manner that does not presuppose familiarity with Hume on the reader's part and yet is sufficiently deep and rigorous to interest more advanced students of his thought. Hume's influence on contemporary epistemology and metaphysics is second to none; probably no other philosopher of the "modern" period continues to have as much influence on the views actually held by contemporary analytic philosophers as does David Hume. Hume is famous, for example, for arguing that meaningful words must have an empirical reference, so that "substance underlying all of a thing's perceivable qualities" and "immaterial soul" lack meaning; that there can be no purely rational demonstration of anything's existence, so that God cannot be proved to exist; that all of our knowledge of scientific laws rests on inferences from experience that are not susceptible of any rational justification; that every claim of the type "A caused B" involves at least one law of nature, so that it is nonsense to talk of cause and effect or causal explanation outside the context of natural science; that the principle that every beginning of existence must have a cause of existence cannot be known *a priori*; that our belief in objects existing independently of our perceptions of them is highly problematic. Such Humean tenets continue to define the broad parameters of much contemporary philosophy, and must be reckoned with by any thinker who wishes to go outside them.

Hume offers his arguments on these and related matters in several key sections of his *A Treatise of Human Nature*, first published in 1739–40, and *An Enquiry Concerning Human Understanding*, published in 1748; those sections are doubtless the Humean texts most frequently encountered by students. One of the chief aims of this book is to analyze and evaluate the arguments contained in those sections, while providing some of the historical background against which they are to be understood. This is worth doing not only because of the intrinsic interest and powerful influence of Hume's arguments, but also because of Hume's manner of exposition. Each work is written in such a way that the reader may be hard pressed to break down Hume's arguments into steps – to analyze them in ways commended by the methods of present-day analytical philosophy. In the youthful *Treatise*,

PREFACE

Hume's style is tortuous and complex; in the mature *Enquiry*, it is so elegant and graceful that the hard muscle and sinews of Hume's thinking may not be apparent. Throughout this book, therefore, one of my chief aims is to present Hume's arguments in a manner which is both sufficiently rigorous to bring out their power and yet sufficiently perspicuous to be accessible to non-specialists in the field. I attempt also to provide informed and reasoned assessments of these arguments; in so doing I discuss some Kantian and several contemporary responses to Hume.

I should also indicate what this book does not attempt to do. The field of Hume studies is today one of the most active in philosophical scholarship, and in the past two decades there have appeared several excellent books proposing new interpretations of Hume's overall system of thought. Some major controversies have emerged: for example whether Hume should be seen as an empiricist and sceptic or as a constructive naturalist. Some novel proposals have been made: for example that Hume believed in objective necessary connections between causes and effects. Although I allude to some of these rival interpretations, that is not my main focus in this book. Rather, this book attempts to explain and assess what I take to be Hume's most *influential* ideas in epistemology and metaphysics. For, perhaps more than any other philosopher of the great age of philosophy that comprised the seventeenth and eighteenth centuries, Hume is a "contemporary" philosopher: his epistemological and metaphysical ideas continue to permeate the current philosophical landscape. It is fitting, then, that an introduction to his epistemology and metaphysics not only guides the reader through the key texts where these ideas are expounded, but also treats them as a living part of ongoing philosophical inquiry.

G. D.
Brockport, New York
July 1997

ACKNOWLEDGMENTS

I am grateful to several anonymous readers for detailed comments on earlier drafts of this book, which led me to make very significant revisions. I am grateful also to the students in my distance-learning Hume–Kant seminar, taught in spring 1997 at SUNY Brockport and transmitted via interactive television to SUNY Cortland and SUNY Fredonia, and to my friend Professor Kenneth G. Lucey, who did me the honor of auditing the course at Fredonia, for comments that led me to clarify several passages. I am solely responsible, of course, for whatever shortcomings remain.

I thank Robert J. McLean, Dean of the School of Letters and Sciences at SUNY Brockport, for support of a course release that enabled me to work on revisions of the manuscript during the Spring 1997 semester.

I am deeply grateful to my companion and colleague Marjorie Stewart for her encouragement and support in all aspects of this book's production.

All quotations of David Hume are from the following editions:

A Treatise of Human Nature, ed. L. A. Selby-Bigge and P. H. Nidditch, Oxford: Oxford University Press, 2nd edn, 1978.

Enquiries Concerning Human Understanding and Concerning the Principles of Morals, ed. L. A. Selby-Bigge and P. H. Nidditch, Oxford: Oxford University Press, 3rd edn, 1975.

All quotations of these two works are included by permission of Oxford University Press. I am grateful to Oxford University Press for permission to reprint this material.

Chapter 2 is based, in part, on my article, "Hume's Fork Revisited," *History of Philosophy Quarterly* 8(4): 327–42. I am grateful to the Editor of *HPQ* for permission to use this material.

NOTE ON REFERENCES AND ABBREVIATIONS

References to the works of David Hume are given in parentheses in the text; references to other authors' works are given according to the Harvard System (author's surname followed by the year of publication of the text cited, and page number(s)). Full bibliographical information is given in the Bibliography.

In the references to and extracts from Hume's *A Treatise of Human Nature*, I have cited the 1978 edition, published by Oxford University Press, and edited by L. A. Selby-Bigge and P. H. Nidditch. References to this work are abbreviated as "T" and are given by page number; for example the reference "(T:188)" is to page 188 of the *Treatise*. Occasionally the citation may be "*T:Abstract*" or simply "*Abstract*," to indicate that the reference is to Hume's *Abstract* of the *Treatise*, which is printed on pages 645–62 of the Selby-Bigge–Nidditch edition.

In the references to and extracts from Hume's *An Enquiry Concerning Human Understanding*, I have cited three editions: the 1975 (Oxford University Press) edition, edited by Selby-Bigge and Nidditch; the 1993 (Hackett Publishing Company) edition, edited by Eric Steinberg; and the 1988 (Open Court Publishing Company) edition, edited by Anthony Flew. The quotations follow the Selby-Bigge–P. H. Nidditch edition; the punctuation and spelling in the other editions may occasionally vary from it. The references to the *Enquiry* cite all three of these editions. The Selby-Bigge–Nidditch edition is cited as "E"; the Steinberg edition is cited as "S"; the Flew edition is cited as "F". For example, the reference "(E:25; S:15; F:71)" is to page 25 in the Selby-Bigge–P. H. Nidditch edition; and the passage cited or quoted may also be found on page 15 of the Steinberg edition and on page 71 of the Flew edition.

The Steinberg and the Flew editions of the *Enquiry* also include Hume's *Abstract of A Treatise of Human Nature*, and the Flew edition includes Section 3 of Book I part iii (on "Why a cause is always necessary") of the *Treatise* as well. Accordingly, when citing the *Abstract* or that section of the *Treatise*, references to these editions are given as well, using "S" followed by the page number for Steinberg and "F" followed by the page number for Flew.

INTRODUCTION

David Hume (1711–76) is chronologically the last, and generally regarded as the most powerful, of the three greatest British Empiricist philosophers, John Locke, George Berkeley, and David Hume. To introduce Hume’s philosophy, a brief survey is made of some of the main intellectual currents that influenced his thought.

First, Hume reacted *against* the metaphysical systems of the Rationalist philosophers, the most important of whom are Descartes, Spinoza, and Leibniz. Each of these thinkers started from certain self-evident principles and attempted to deduce from these principles a complete system of knowledge. Following this method, however, they arrived at radically different conclusions. Thus Descartes, starting from his famous “I am thinking, therefore I exist” and a small number of other principles that he found to be self-evident – such as the principle that a cause must contain as much perfection as its effect, and the principle that whatever things we can conceive clearly and distinctly to exist separately from each other can really exist separately from each other – tried to demonstrate that God exists and that the universe God created consists of two completely different kinds of substance: minds, which are totally nonphysical and whose only properties are their thoughts or conscious states; and matter, which is totally incapable of thought or consciousness, and whose defining property is extension (i.e. three-dimensionality). But Spinoza, starting from various “axioms,” “definitions,” and “postulates” that he took to be self-evident, arrived at the conclusion that the whole universe consists of only one infinite substance which is both conscious and three-dimensional, which may be called either “God” or “Nature” depending on one’s point of view, and of which individual persons and things are only finite aspects or “modes.” Finally Leibniz, starting from several principles with names like the “Principle of Sufficient Reason,” the “Principle of Perfection,” and “the Identity of Indiscernibles,” arrived at the conclusion that the universe consists of an infinite number of nonmaterial substances, called “monads,” which are differentiated only by the perceptions they have, and that extension is not a real property of those monads but one that they only appear to have.

What characterized each of these Rationalist philosophers, then, was this: they speculated boldly and confidently about the ultimate nature of reality, and their confidence in the conclusions they reached stemmed from their conviction that those conclusions were rigorously

INTRODUCTION

deduced from self-evident premisses. Hume, however, saw that by following essentially the same method of deducing conclusions from supposedly self-evident premisses these mighty thinkers had arrived at drastically different, mutually incompatible, results. His reaction, as we shall see later, was to criticize the theory of knowledge that underlies the Rationalists' method, and to propose a very different theory of knowledge to replace it.

A second, more positive, influence on Hume was Isaac Newton, the founder of classical physics whom Hume admired and even revered. Newton did not develop his physics by arguing deductively from supposedly self-evident premisses. Rather, he confined himself to hypotheses that could be experimentally tested, thereby shedding enormous light on the workings of nature. Hume, as we shall see, sought to adapt Newton's experimental method to his own inquiries.

A third influence on Hume was John Locke, the founder of the British Empiricist school. Three aspects of Locke's thought are especially relevant. The first is what we may call Locke's "epistemological turn." This is the view that before tackling big questions about the nature of reality – such as the existence and nature of God, or the basic properties of matter, or the immortality of the soul – we need to investigate the human mind with a view to ascertaining both its powers and its limitations, so that we are enabled to determine what we may realistically hope to know. The second influential aspect is what Locke called the "Way of Ideas." This is the view, derived mainly from Descartes, that what each human knows best and with certainty is his or her own conscious states or "ideas," and that all the rest of one's knowledge must in some fashion be based upon these ideas. Thus, for example, at this moment you presumably know that there are several physical objects close to you, such as the book you are reading, the desk or table at which you may be sitting, the windows and walls of the room in which you are, and so forth. According to the way-of-ideas doctrine, this knowledge must be based on certain conscious states you are in, such as visual experiences of color and shape and tactile sensations of hardness or solidity. Furthermore, since you could have similar experiences in a vivid dream or hallucination, the manner in which your knowledge is based on those conscious states is problematic and indeed calls for an explanation or even for a theory. The third aspect is Locke's famous denial of innate ideas. Briefly, an innate idea would be one that had not been acquired or extrapolated from any experience one had, because the idea was somehow possessed by or built into the mind from birth. Hume, as we shall see, agrees with Locke that all of our ideas must originate in experience.

As we shall see later, there were still other intellectual currents that influenced Hume's thinking, notably philosophical scepticism.¹ But against the background of the three factors just surveyed, it is possible to describe in general terms Hume's basic "agenda" or program.² Hume sought to adapt the experimental method of Newton to the investigation of the powers and principles of the human mind launched by Locke. Here I have said "adapt" rather than "adopt," because Hume did not think that physical experiments could be performed on the mind. Rather, he thought that the mind's workings are accessible to introspection, and

INTRODUCTION

that by careful introspective study of one's own conscious states, one would be able to discover general principles that apply to those states; much as by carefully studying the operations of physical objects Newton had discovered general principles applying to them, such as the laws of motion and gravitation. The result of this essentially introspective study of the mind was to be a truly empirical science of human nature. Hume would then use the findings of this new science of human nature, negatively, to criticize the overly ambitious theories of rationalist metaphysicians. He would also use his findings, positively, to offer his own accounts of the origin of certain basic human beliefs: for example, the belief in causal connections between events; the belief in the existence of objects independently of our perceptions of them; and the belief in the existence of a continuing mind or self.

Hume's first philosophical work was his *A Treatise of Human Nature*, an enormous book of over 600 pages that he published anonymously when he was only 28 years old, after several years of intense labor that left him drained and in ill health. Probably because of its difficult style, great length, and revolutionary content, the *Treatise* was initially a failure. As Hume put it in his brief autobiography "My own Life" (1993: 351–6): "Never literary attempt was more unfortunate than my *Treatise of Human Nature*. It fell dead-born from the press, without reaching such distinction as even to excite a murmur among the zealots" (p. 352).

Hume recovered from his disappointment, and wrote two further works in which he presented many of the *Treatise's* themes in a more accessible literary manner: *An Enquiry Concerning Human Understanding* and *An Enquiry Concerning the Principles of Morals*. Although the two *Enquiries* were better received than had been the *Treatise*, Hume's fame during his own lifetime was based primarily on his non-philosophical writings. These include a six-volume history of England and several political and literary essays. Hume's last philosophical work was his *Dialogues Concerning Natural Religion*, a classic critique of arguments for the existence of God that was published only after his death.

Our focus will be primarily on certain key sections of *An Enquiry Concerning Human Understanding* (hereafter called simply "the *Enquiry*") and of Book I of the *Treatise*. Occasionally I shall refer also to a short work called *An Abstract of a Book Lately Published, Entitled, A Treatise of Human Nature, &c* (hereafter called simply "Abstract"). This is actually a book review of the *Treatise*, written anonymously by Hume himself, which contains clear brief expositions of several of the points on which our attention will focus.

HUME'S THEORY OF MEANING AND ITS IMPLICATIONS

1 Hume's theory of impressions and ideas and meaning-empiricism

Both the *Treatise* and the *Enquiry* open, after introductory sections, with almost identically titled chapters: "Of the Origin of our Ideas" and "Of the Origin of Ideas", respectively. As the titles indicate, the purpose of these sections is to explain how we acquire our ideas. Hume's basic thesis, which is a cornerstone of his empiricism and which had already been enunciated by Locke in his *Essay Concerning Human Understanding*, is that we acquire all of our ideas from experience, where "experience" is taken as including both sense perception and the introspective awareness of our own states of mind. From this starting point, Hume will ultimately derive momentous consequences.

In setting forth his basic thesis, Hume uses three special terms: "perception," "impression," and "idea." By a *perception*, Hume means any conscious state whatsoever. Hume divides perceptions into two classes: impressions and ideas. By an *impression*, he means any experience, such as a visual experience, an auditory experience, or a pain. Hume further subdivides impressions into two classes: sense experiences, such as visual, tactual, auditory, gustatory, olfactory, and kinesthetic experiences; and "inner" or introspectible experiences, such as joy, sadness, anger, and desire. In the *Enquiry*, he refers to impressions of the first kind as "outward sentiments" or "outward sensations" and to impressions of the second kind as "inward sentiments" or "inner sensations" (E:19, 22; S:11, 13; F:65, 67). In the *Treatise*, on the other hand, he uses a more technical terminology that is borrowed in part from Locke: he calls impressions of the first kind "impressions of sensation" and impressions of the second kind "impressions of reflection" (T:7). Here the term "reflection" does not mean that such impressions are more abstract or intellectual than are impressions of sensation, but rather that one becomes aware of them by introspection – by a kind of reflection on one's own state of mind. By an *idea*, Hume means any conscious state other than an impression. Ideas include especially concepts, but also mental images, such as those that occur when one imagines something or, sometimes, when one remembers something. In the *Treatise*, Hume

also observes that the introspectible experiences, or impressions of reflection, are often caused by ideas; for example, an idea of pain may cause an impression of fear, or an idea of pleasure may cause an impression of desire (T:8).

The distinction between “impressions” and “ideas” probably seems fairly obvious to you. Hume himself thought it quite unproblematic, saying that “it requires no nice discernment or metaphysical head to mark the distinction between them” (E:18; S:10; F:63). Before seeing how Hume uses these terms in formulating his basic thesis, however, we need to take note of an ambiguity in his account of the difference between impressions of sensation and ideas. Hume’s main or “official” way of distinguishing between impressions and ideas is to say that impressions are lively, vivid, or “forcible,” whereas ideas are “faint” or “dull.” But Hume adds immediately that a person whose mind is “disordered by disease or madness” may have ideas that are “altogether indistinguishable” from his or her impressions (E:17; S:10; F:63). Hume’s point here is that people who are insane, or suffering from psychotic delusions or hallucinations, may have ideas which are every bit as vivid or lively as their impressions of sensation. Now in saying this, Hume is tacitly appealing to a criterion for distinguishing between impressions of sensation and ideas other than vivacity, because he still calls a deranged person’s vivid hallucinations “ideas;” but their vividness would force him to classify them as impressions, if vividness or vivacity were the operative criterion for distinguishing between impressions and ideas. What other criterion, then, is Hume appealing to? As Jonathan Bennett has pointed out (1971: 224–5), he is appealing to the criterion of *objectivity*, according to which impressions of sensation are experiences had when people really perceive physical objects; whereas hallucinations, no matter how vivid they may be, are only ideas. There are several other places where Hume does this: in the *Abstract*, he says that “when we . . . have the images of external objects conveyed by our senses, the perception of the mind is . . . an *impression*,” and he equates the claim that “our [visual] ideas [of sensation] . . . are derived from impressions” with the claim that “we can never think of anything which we have not seen without [outside] us” (T:647; S:128; F:31); in the text of the *Treatise* proper, he contrasts “that idea of red, which we form in the dark,” with “that impression, which strikes our eyes in sun-shine” (T:3); in the *Enquiry*, he introduces the term “impression” by listing “perceptions, when we hear, or see, or feel” among the examples of what he will mean by that term (E:18; S:10; F:64). In these passages, from which we have quoted only short excerpts, Hume *also* characterizes impressions as being our “more lively” or “strong” perceptions, but the fact remains that he equates impressions of sensation with those had when we are really perceiving physical objects or, in Bennett’s language, having “experience of the objective realm” (1971: 224–5). Thus, Hume has two different and incompatible criteria for distinguishing between impressions of sensation and ideas: his official criterion of “force and vivacity,” and the implicit and unacknowledged criterion of objectivity.

The probable explanation of this ambiguity stems from a factor that we mentioned (but postponed describing) in the Introduction, when surveying the intellectual currents that

influenced Hume: namely, his preoccupation with philosophical scepticism. Scepticism, a philosophical position that dates back to ancient Greece and has proponents to this very day, calls into question the possibility of knowledge. It does so by using certain arguments that are designed to show that our cognitive faculties and powers – our senses, reasoning ability, and memory – are not reliable enough to enable us to distinguish securely between truth and error, appearance and reality, well-grounded belief and mere opinion. The type of scepticism that is directly relevant to Hume’s conflicting criteria for distinguishing between impressions and ideas is scepticism about the reliability of the senses. Such scepticism was a major force in Hume’s day, and was advocated by one of Hume’s favorite writers, Pierre Bayle. But to understand it, we should refer to Descartes, one of its most famous and powerful sources.

Descartes himself was no sceptic; indeed, he attempted to refute scepticism. But in preparing the way for his refutation, he first presented the case *for* scepticism so as to be in a better position to demolish that case. Indeed, his strategy was to radicalize the case for scepticism – to make it more sweeping and powerful than the sceptics themselves had done – and then to refute the radicalized sceptical arguments so as to abolish scepticism once and for all. Descartes begins modestly enough, by pointing out that our senses are sometimes deceptive, as when we misperceive an object’s shape or size.¹ Of course, he then notes, this hardly shows that our senses are never trustworthy, since such errors typically occur because the object is far away or seen through a fog or the like. But why can’t we rely on our senses when the conditions of observation are favorable, for example, when viewing a nearby object in good light? Descartes answers with two different sceptical arguments. First, he argues, as had other philosophers before him, that even our very best sense perceptions – those that occur under the most favorable conditions of observation – can be duplicated in vivid dreams. In other words, some dreams are so life-like, so vivid, that we are completely taken in by them and so cannot reliably distinguish them from our best waking perceptions. Second, he introduces an argument that no one before him had used and that is now commonly regarded as his most important sceptical argument. This is the argument that perhaps, for all we can tell, all of our sense experiences are produced by a source completely different from that which we think is responsible. We think, of course, that our visual, tactile, auditory, gustatory, and olfactory experiences are produced when physical objects stimulate our sense-receptor organs – our eyes, skin, ears, taste buds, noses. But what if, instead, they are caused in some radically different way; say by God himself, or perhaps by some powerful and evil demon bent on deceiving us? At first, this argument may seem utterly bizarre, perhaps even laughable. But on reflection it tends to “grow on” one. For, after all, a very powerful and intelligent being such as God, or some evil counterpart of God, presumably could generate in humans a perfectly undetectable hallucination of an entire physical world. All that this being would have to do is to cause us to have sense experiences that are vivid and that fall into orderly coherent patterns, so that we could, for example, correctly and consistently predict future experiences on the basis of present ones. How can we possibly

know that this is not the way our experiences are in fact produced, since by hypothesis the experiences themselves would be exactly the same if they were so produced?

Descartes thought that he could refute this “deceiver argument.” He attempted to refute it by offering proofs that there exists a benevolent God who would not deceive his creatures about the causes of their sense experiences. But Descartes’ proofs of God’s existence convinced few philosophers, so that his presentation of the problem of scepticism has endured while his attempt to solve it has not. Other philosophers have attempted to provide more effective responses to scepticism. John Locke, for example, attempts to solve Descartes’ problem without appealing to God, by means of a quasi-scientific argument to the effect that the best explanation of the order and coherence of our sense experiences is that they are produced by a world of physical objects interacting with each other and with our sensory organs in regular, lawful ways. But Locke’s successor in the British Empiricist tradition, Bishop Berkeley, argued powerfully that nothing Locke said successfully rules out other possible explanations of the order and coherence of our sense experiences. He then proceeded to argue that the experiences are indeed caused by God himself; except that Berkeley saw nothing sceptical about this view, because he combined it with the extraordinary thesis that matter does not really exist, and that what we call physical objects really are nothing but orderly groups of sensations or ideas caused by God!

Hume fully appreciated the force of Descartes’ sceptical arguments, and he agreed with Berkeley’s criticisms of Locke. But Hume could not accept Berkeley’s own cure for scepticism, and indeed regarded Berkeley as just another sceptic. Accordingly, Hume adopted a thoroughgoing scepticism concerning the senses: he maintained that our belief in a material world, though it is a belief that we cannot avoid having, is not susceptible to any rational justification.² In the *Treatise*, he says that our impressions of sensations arise “from unknown causes” (T:7). So, Hume tries to find a criterion for distinguishing impressions of sensation from ideas that does not depend on knowing whether any of the impressions are caused by physical things – a purely immanent or phenomenological criterion that can be applied simply by using introspection. The criterion he comes up with is vivacity or liveliness; and this he adopts as his official way of distinguishing between impressions and ideas.

However, when Hume is not actually discussing the problem of perception, he tends not to keep in mind his scepticism concerning the senses. For example, as we shall see in Chapters 3 and 4, Hume’s discussions of cause-and-effect reasoning, and of the nature of the cause-effect relationship itself, are among his most important contributions to philosophy. But as we shall also see, his discussion of those topics is written from a point of view that just takes for granted our knowledge through perception of physical objects and events; while discussing causal reasoning and causality, his deep interest in these topics simply overrides his sceptical doubts about the senses. This is not necessarily an inconsistency on Hume’s part; for he can be interpreted as asking the questions: *if* we had knowledge of physical objects and events, then what would be the correct accounts of causal reasoning about them and of causal relationships among them? In any case, it is fortunate that Hume

does abstract from or put to one side his scepticism about the senses while discussing causal reasoning and causality, because it is doubtful that he could have made his seminal claims about these topics if he had consistently stuck to such a sceptical stance, allowing himself to speak about impressions and ideas only and never about physical objects.

Now another place where Hume does not (or, at any rate, not always) bear this scepticism in mind is in presenting his theory of impressions and ideas. This is why he allows himself to appeal to objectivity as a criterion for distinguishing impressions of sensation from ideas. But, in this instance again, the resulting ambiguity, while potentially confusing, turns out not to be damaging. For, as we shall see, the importance of Hume's theory of impressions and ideas lies chiefly in the way he uses the theory to criticize certain conceptions taken for granted by his philosophical predecessors, such as the conception of substance, and his use of the theory does not depend on which of the two criteria is used to distinguish between impressions and ideas.

Accordingly, let us grant Hume his distinction between impressions and ideas, without worrying further about the ambiguity just discussed. We can then set forth Hume's basic principle about the relationship between impressions and ideas, which is his main opening thesis in both the *Enquiry* and the *Treatise*. This principle, which we will call the "principle of empiricism," can be put as follows: every idea is either

- (a) derived from a corresponding impression, or
- (b) composed of simpler ideas, each of which is derived from a corresponding impression.

Here "derived from" means, roughly, *copied from* and "corresponding" means *resembling*. An example of an idea falling under (a) would be that of some specific shade of red, say crimson. Hume would say that this idea is derived from a corresponding impression – an experience in which crimson itself is presented. An example of an idea falling under (b) might be the idea of a centaur. For, suppose a person has never seen a centaur (since there are none), nor ever had a vivid hallucination of one (so as to guarantee that the person has never had an impression of a centaur, whichever criterion we use for distinguishing impressions from ideas). Still, that person can have an idea of a centaur. For this idea can be broken down into two simpler components, such as the idea of a human being's head and torso and the idea of a horse's trunk and limbs. So, provided the person has had impressions of people's heads and torsos and of horses' trunks and limbs (impressions that can be obtained simply by seeing people and seeing horses), he or she can have the idea of a centaur. For then the person can have the ideas of a human head and torso and of a horse's trunk and limbs, and can combine these ideas to form the idea of a centaur. In a similar way, a person can have innumerable ideas that have no exact counterparts among impressions, provided that those ideas can be analyzed into component ideas all of which do have counterparts among the

impressions. As examples of such ideas, Hume mentions those of a golden mountain and a virtuous horse.

Given the importance that Hume attaches to his principle of empiricism, his defense of the principle is surprisingly brief and also vulnerable to attack. First, he simply challenges the reader to find a counterexample to the principle; that is, to find an idea – any idea – that is not derived from impressions in either manner (a) or manner (b). Second, he claims that if a person's sensory organs are defective, so that the person cannot obtain the impressions of sensation normally generated when those organs are stimulated, then he or she cannot have the corresponding ideas: "A blind man can form no notion of colours; a deaf man of sounds" (E:20; S:12; F:65).

At first, these two points may seem quite persuasive; but on further reflection both become quite problematic. To begin with the first point, it is somewhat surprising that Hume himself, shortly after challenging the reader to find an idea which is not derived from any impression, describes a case that meets this very challenge! This is his famous case of the "missing shade of blue." Hume asks us to imagine a person who has seen every shade of blue from the darkest to the lightest, except for just one intermediate shade. Hume then asks whether this person could conjure up imaginatively the idea of that "missing" shade of blue, despite never having had an impression of it. Hume answers, plausibly enough, that the person could conjure up or visualize the missing shade. Notice that this case is a direct and convincing counterexample to Hume's own principle. For the idea of the missing shade of blue is by hypothesis not derived from any corresponding impression; nor can it be broken down into simpler ideas each of which is derived from a corresponding impression, because it is not composed of parts. Rather, the idea of any shade of blue is uniform, uncompounded, or simple; so that, unlike the idea of a centaur, it cannot be divided up into component ideas, each of which might be derived from different impressions. Furthermore, it seems that other examples of the same sort can easily be given, involving shades of other colors and perhaps also sounds. Yet, Hume dismisses the example as unimportant, saying that "it does not merit, that for it alone we should alter our general maxim" (E II:21; S:13; F:66; T I i 1:6). This is certainly puzzling: how can Hume maintain his principle, even though he knows that there is a clear counterexample to it?

Hume's second point – about defective sensory organs – also runs into difficulty. To see why, consider a hypothetical case. Suppose that instead of becoming deaf in his later years, Beethoven had been deaf from birth, and so had never had any impressions of sound. What could Hume say if our congenitally deaf Beethoven assured him that he knew very well what sound was, and presented him with the scores of his nine symphonies, thirty-two piano sonatas, etc., as evidence? It looks very much as if, in the face of such evidence, Hume would have to admit that his principle was false. Here you may be inclined to defend Hume by saying that the case of a congenitally deaf Beethoven does not really count against Hume's principle, because it is just an imaginary example. In order to show that Hume's principle is

false, one must show that there really is a person who has ideas without having had the corresponding impressions.

The trouble with this possible defense of Hume, however, is that it does not square with the way he himself treats his principle, for it treats the principle as a falsifiable empirical or a posteriori thesis, whereas Hume treats it as an unfalsifiable a priori thesis. To explain this point, we need to define the terminology used. An empirical or a posteriori statement is one that can be known only by experience; for example, the statements that snow is white, or that water boils at 212°F at sea level, or that thunder follows lightning. An a priori statement is one that can be known just by thinking; for example, the statements that $1 + 2 = 3$, or that nobody is his or her own mother, or that if X is to the left of Y , then Y is to the right of X . We shall meet this basic distinction again, and discuss it in more depth, in the next chapter. For the present, we need to note only one more point about it. This is that empirical generalizations, which are statements of the form “All A s are B s,” are always falsifiable; meaning that one can always conceive of some experience or observation that would show such a statement to be false. For instance, finding a patch of purple snow would show that the statement *all snow is white* is false; finding some water that failed to boil when heated to 212°F at sea level would falsify (show to be false) the statement *all water boils at 212°F at sea level*, and so forth. (Incidentally, notice that calling a *statement falsifiable* is, of course, not the same as calling it false: it means that some possible experience *could* conceivably show the statement to be false, not that the statement actually is false; so that many statements of whose truth we are quite certain are nevertheless falsifiable.) By contrast, no a priori statement is falsifiable. We cannot conceive, for example, of any experience that would show $1 + 2 = 3$ to be false. For suppose that we try to do so: imagine, for example, inserting two marbles through a small opening to a box already containing one marble, and that on opening the box you find only two marbles therein. Would that falsify the arithmetical statement $1 + 2 = 3$? Or would it show that $1 + 2 = 2$? Surely not. At best, it would show that one marble had somehow exited the box while you were not looking. Failing some such natural explanation, it might even show that one of the marbles had somehow dematerialized; but it would not show $1 + 2 = 3$ to be false.

To return to Hume’s principle. We saw that this principle can be challenged by the hypothetical case of a congenitally deaf Beethoven. We said that a possible reply to that case is that it is merely imaginary, and that Hume’s principle can be refuted only by a real case. Now this possible reply obviously assumes that the principle is falsifiable, and therefore also that it is an empirical generalization to the effect that all ideas are, as a matter of ascertainable fact, derived from corresponding impressions or else are composed of simpler ideas each of which is so derived. But although Hume sometimes writes (especially in *Treatise* Book I, part i, Section 1) as if his principle were an empirical hypothesis, we shall see that he does not treat it that way. For every time a case arises where people claim to have an idea for which Hume can find no impression or impressions (as in the cases of the ideas of substance and of self, to be discussed in the following part), he concludes that people do

not really have the idea in question (and usually proposes some substitute idea that can be derived from impressions). He does not consider seriously the possibility that his principle might instead be false; even when, as in his “Appendix” to the *Treatise* (T:633–6), he sees that it has led him to a view of the self that he himself finds untenable. Thus, despite the fact that Hume sometimes speaks as if his principle were a falsifiable empirical generalization, in practice he treats it rather as an a priori truth. But then his appeal to empirical evidence to support the principle – for instance, to the fact that people with defective sensory-organs lack the corresponding ideas – is irrelevant and misleading, since the principle is not really based on empirical evidence. Furthermore, it is not at all clear how a generalization about the origin of our ideas *could* be known a priori.

In what follows, we shall propose a way of interpreting Hume's principle that solves the difficulty raised by the missing shade of blue and the difficulty raised by the congenitally deaf Beethoven scenario. Let us begin by considering the final paragraph of the second section of the *Enquiry*. For this paragraph sets Hume's principle in a certain light, in which the cases of the missing shade of blue and of a congenitally deaf Beethoven no longer cause trouble for the principle. The paragraph reads as follows:

Here, therefore, is a proposition, which not only seems, in itself, simple and intelligible; but, if a proper use were made of it, might render every dispute equally intelligible, and banish all that jargon, which has so long taken possession of metaphysical reasonings, and drawn disgrace upon them. All ideas, especially abstract ones, are naturally faint and obscure: the mind has but a slender hold of them: they are apt to be confounded with other resembling ideas; and when we have often employed any term, though without a distinct meaning, we are apt to imagine it has a determinate idea annexed to it. On the contrary, all impressions, that is, all sensations, either outward or inward, are strong and vivid: the limits between them are more exactly determined: nor is it easy to fall into any error or mistake with regard to them. When we entertain, therefore, any suspicion that a philosophical term is employed without any meaning or idea (as is but too frequent), we need but enquire, *from what impression is that supposed idea derived?* And if it be impossible to assign any, this will serve to confirm our suspicion.

(E:22; S:13; F:67; see also *Abstract* in: T:649; S:129; F:32)

In this passage, Hume introduces a topic that he has not previously mentioned: the topic of meaning. He alludes to “jargon” (meaningless verbiage, gibberish), and implies that certain disputes are unintelligible. He twice mentions the possibility of a term being employed without any meaning, and tells us how to guard against this. Indeed, what Hume does in this paragraph is to derive from his principle of empiricism a test for meaning. This test can be put as follows:

If T is a general, classificatory, or descriptive term that supposedly has an idea, *I*, as its meaning, but there is no impression(s) from which *I* is derived, then T does not stand for any idea and is therefore meaningless.

The last two sentences of the passage quoted foreshadow the way in which Hume will apply this test. He will apply it to three philosophical terms that are central in the work of nearly all his predecessors: the terms “substance,” “self,” and “cause.” Indeed, as we shall see, some of Hume’s chief contributions to philosophy arise when he subjects these three terms to his test for meaning.

Today, the importance of Hume’s principle of empiricism is generally taken to lie primarily in the test for meaning that he derived from it. Indeed, recent Empiricist philosophers (who still look to Hume as their philosophical progenitor) are not generally satisfied with Hume’s formulation of the principle of empiricism. For Hume’s principle is, as we have seen, a genetic principle – one about the origin or genesis of our ideas. As such, it runs into the difficulties that we have noted: the problem that a person might, for all we know, have ideas that did not originate in any impressions (e.g. our congenitally deaf Beethoven’s ideas of sound), and the problem that a person who had received impressions of all but a few shades of a color could surely visualize the shade(s) that he or she had never experienced (e.g. the missing shade of blue). Furthermore, it seems implausible to maintain that whether or not a term has a meaning depends on whether a given person, or several persons, have had impressions answering to (the idea associated with) that term. On the other hand, it is not implausible to maintain that whether or not a descriptive term has a meaning depends on whether something that *could* be experienced would answer to that term (or at least to other terms by which it was defined).

Accordingly, most contemporary Empiricists would recast Hume’s entire position as one whose concern is about meaning. Thus, they would recast his principle of empiricism along the following lines, suggested by H. H. Price:

Every meaningful general term T is definable either: (a’) ostensively, i.e. by indicating something that can be encountered in experience and saying: “that is an instance of what I mean by ‘T’”, or (b’) by means of other terms that are themselves definable in manner (a’).

(Price 1965: 6)

There is a very close parallel between this “Empiricist view of meaning” (as we may call it), and Hume’s principle of empiricism. For, notice that we can very plausibly use the same example, “crimson,” to illustrate (a’) that we used to illustrate (a), and the same example, “centaur,” to illustrate (b’) that we used to illustrate (b). Furthermore, from the Empiricist view of meaning, we can derive a modernized version of Hume’s test for meaning: if T is a general or classificatory or descriptive term, but T cannot be defined either in manner (a’) or in manner (b’), then T is a meaningless term.

The main difference between Hume's principle of empiricism and test for meaning and their modern counterparts is this: in Hume's formulations, considerations of meaning are intertwined with a psychological, genetic thesis about the origin of our ideas – about how we *learn* meanings. The modern formulations, by contrast, are purely principles about what it is for a term to *have* meaning. As suggested above, this gives the modern formulations an advantage: they do not face the difficulties that afflict Hume's own position. Consider first the case of the missing shade of blue. As we have seen, this case is a clear counterexample to Hume's formulation of the principle of empiricism. But it is not a counterexample to the Empiricist view of meaning. For that view requires only that some observable or experienceable item *would answer* to the term for that shade of blue. It does not require anyone to have actually seen that shade. It requires only that the term's meaning *could* be defined ostensively or, as Price (1965: 7) put it, that the term be "cashable" in experience. A similar point solves the difficulty raised by the case of a congenitally deaf Beethoven. If Hume's thesis about the relation between impressions and ideas is construed along the lines suggested by the Empiricist view of meaning, then it requires only that every idea could be *exemplified* by an impression (or be composed of simpler ideas that could be exemplified by impressions), not that it be derived from an impression. But if someone had ideas that were not *derived* from any impressions, it would not follow that those ideas could not be *exemplified* by impressions, that they could not be "cashed" in experience. Even if Beethoven had really been congenitally deaf and despite this still had ideas of sounds, it would not follow that those ideas could not be exemplified by impressions. On the contrary, they could be exemplified by impressions of sound, despite Beethoven's never having had such impressions. Those ideas would still be "cashable" in experience, even though they had not *originated* in experience. To put the point differently: even if Beethoven had been congenitally deaf and nonetheless had ideas of sounds, it would not follow that nothing in any experience he could have had, had he not been congenitally deaf, would have answered to those ideas. On the contrary, auditory impressions, which he then could have had, would have answered to his ideas of sound.

Following Bennett (1971: 225–30), let us call the modernized versions of Hume's principle of empiricism and test for meaning "meaning-empiricism." Meaning-empiricism, as we have just seen, avoids the problems that we raised for Hume's own formulations.³ Does this mean that we should accept meaning-empiricism? Not necessarily; for it may be that there are terms which are meaningful, but do not satisfy the Empiricist conditions for meaning. Hume himself, as we mentioned, examines three terms that his predecessors took to be not only meaningful, but also central to our thinking: "substance," "self," and "cause." In the next part, we shall consider what he says about substance and self, leaving "cause" aside until we have set out his related views about causal reasoning. At the end, we should be in a better position to assess meaning-empiricism.

2 Application of Hume's test for meaning to "substance" and "self"

In this part, we shall apply the Empiricist test for meaning to "substance" and "self." In order to prepare the ground, however, we need to recount how the notions of substance and self were understood by the philosophers whose views Hume was criticizing.

The notion of *substance* is the centerpiece of a view that we shall call the "substance theory." The substance theory, which can be traced back at least to Aristotle and which was upheld by most medieval thinkers and by the Rationalists of the seventeenth and eighteenth centuries, is essentially an attempt to answer the philosophical question: "What is a thing?" There are two competing answers to this question: the bundle theory and the substance theory. According to the bundle theory, which was favored by Empiricist philosophers like Berkeley, Hume himself, and (in the twentieth century) Bertrand Russell, a thing is nothing but a collection of coexisting properties. For example, a tomato is nothing but roundness, redness, squashiness, juiciness, and so on, existing together at a certain place and time. According to the substance theory, however, a thing is composed of more than just its properties: it is composed of those properties *plus* an underlying substance to which all the properties belong.⁴ Thus the tomato, on this view, is composed only in part of the properties just mentioned; for it is also composed of an underlying substance to which all those properties belong and in which they are said to "inhere." As a very rough analogy, think of a pin-cushion with pins stuck in it: the pins are analogous to the properties, the pin-cushion to the substance.

On the standard seventeenth- and eighteenth-century view stemming largely from Descartes, there are two different kinds of substance. One is matter or material substance, whose basic properties are shape, size, and solidity.⁵ The other is mind or mental substance, whose *only* properties are conscious states. Furthermore, according to this view, which is called "substance dualism," "Cartesian dualism," or sometimes simply "dualism," a person or self is identified with his or her mind or mental substance.

The principal rationale for the substance theory is provided by what is called the "argument from change." This argument is implicit in a very famous passage in the *Second Meditation*, where Descartes describes what happens to a piece of wax, freshly taken from a beehive, when that wax is put near a fire. As the wax is heated, its properties change: its hardness is replaced by a soft, gooey texture, its lumpish form by an elongated shape, its brown color by a translucent tint, its fragrant aroma by a smoky smell. Even its capacity to make a noise when struck with a finger is lost. Yet, one and the same piece of wax still exists despite all these changes. Why is that? Why isn't it the case, instead, that the wax ceases to exist and that another and new object begins to exist? The answer, according to the argument from change, is that although the properties of the wax have changed, the underlying substance has not: one and the same substance existed throughout the process of change and still exists now.

This is an important argument, so let us formulate it carefully. To do this, we need to make a distinction between properties of two kinds: determinate properties and determinable properties. A determinate property is one that is absolutely specific, whereas a determinable property is one that is not absolutely specific. Take for example the property of being elliptical, or ellipticalness. This is a determinable property, not a determinate one, because there are many different elliptical shapes: thick ones, thin ones, in-between ones, and so forth. A determinate property, by contrast, must be absolutely specific. Only an elliptical shape satisfying a particular mathematical equation (ellipticity) would be a determinate property. Color is another example of a determinable property, since there are many different colors. Even particular colors like red, blue and yellow are determinable rather than determinate properties, because there are many shades of each of those colors. Only absolutely specific shades would be determinate properties, even if we do not have names for each such shade.

Having distinguished between determinate and determinable properties, we can now state the argument from change accurately:

- (1) We can distinguish between (a) all of a thing's determinate properties changing without the thing's ceasing to exist and (b) a thing's ceasing to exist.
- (2) We can distinguish between (a) and (b) only if a thing is composed, in addition to its properties, of a permanent underlying substance.

∴ A thing is composed, in addition to its properties, of a permanent, underlying substance.

To grasp this argument, think again of Descartes' example of the melting wax. The example can be understood as raising a challenge: namely, what justifies us in taking the melting wax to be a case of (a) rather than a case of (b) – that is, in thinking that the wax, all the determinate properties of which have changed, has not therefore ceased to exist? Premiss (2) says, in effect, that the only way to answer this challenge is to admit that the wax is composed of something *more* than just its properties, of something permanent to which those properties belong. The conclusion is simply a statement of the substance theory itself, so that the argument as a whole purports to prove that this theory is correct.

You may be wondering why the term “determinate” is needed in premiss (1). The answer has to do with the nature of change. We say, of course, that when a thing alters, its properties or characteristics have changed. But what does this mean? Not that the properties *themselves* have changed, because a property itself can never change: red, for example, is just red, and to say that it had changed would really mean that it had been *replaced* by another property, say by the property blue. Thus, to say that a thing's properties have changed is to say that it has lost certain properties and acquired others. Suppose, then, that the word “determinate” were left out of premiss (1). Then the premiss would say that a thing could lose all of its properties, *including its determinable properties*, without ceasing to exist. But this would be

false, for a thing's determinable properties include shape and size – not this or that specific shape and size, but just having some shape and size or other, or having what Descartes called “extension” – and a (physical) thing cannot lose all shape and size whatsoever without thereby also ceasing to exist. Thus, the term “determinate” must be included in premiss (1) for that premiss to have a chance of being true.

You may now ask: why then isn't the word “determinate” also included in premiss (2)? The answer is that to preserve the argument's validity, the conclusion would then also have to include that word, and so would have to read: “A thing is composed, in addition to its determinate properties, of a permanent underlying substance.” But this is false, for it means that a thing is composed *only* of its determinate properties and its underlying substance. But a determinate property cannot possibly be present unless a corresponding determinable property is also present; for example, squareness cannot be present unless shape is present. Thus, in order for the argument's conclusion to stand a chance of being true, the word “determinate” must not be included in premiss (2).

The argument from change can also be applied to the non-physical, purely mental, thing that most seventeenth- and eighteenth-century philosophers took a human mind to be. Thus Descartes wrote:

[T]he human mind is a pure substance. For even if all the accidents of the mind change, so that it has different objects of the understanding and different desires and sensations, it does not on that account become a different mind.

(1984: 10)

Here the term “accidents” refers to properties of a certain kind – namely, accidental properties. Accidental properties are those that a thing may have but need not have in order to be what it is, such as a certain triangle's property of being green. Accidental properties contrast with essential properties, which are those that a thing must have in order to be what it is, like the triangle's property of being three-sided. Descartes' phrase, “different objects of the understanding and different desires and sensations,” here refers to a mind's specific, determinate conscious states. Thus, Descartes identifies the accidental properties of a mind with its determinate, conscious states. But if a mind's accidental properties are determinate conscious states, then those properties are also determinate properties. Therefore, the argument of the passage could be summarized: “A human mind is a substance, because even if all its determinate properties change, it is still the same mind.” This is simply the argument from change applied to the mind rather than to a physical thing like the wax.

Although the substance theory was upheld by most major philosophers of Hume's day, it is not without its problems. Before looking at Hume's own specific objection to the theory, it will be useful to state briefly the central difficulty faced by the theory. This is that substance is unperceivable. Imagine, for example, that you wanted to see the substance of a

piece of wood. So you obtain a carpenter's plane and shave off a thin layer of wood. What do you then see? Well, you see a new set of properties – a slightly diminished size, a slightly altered shape, a slightly different shade of color. It is obvious that planing away still more layers of wood will not get you any closer to seeing the piece of wood's underlying substance. No matter how many layers you plane away, you will see only more properties – until, at last, all the wood is gone. What this type of thought experiment reveals is that substance is just not something that can be perceived: nothing one could do would even count as perceiving substance.⁶ Substance is, as philosophers say, unperceivable *in principle*. For this reason, some philosophers of the seventeenth and eighteenth centuries became increasingly suspicious of the notion of substance; and many contemporary philosophers reject “substance” altogether.

Against the background of the substance theory and the fundamental difficulty that it faces, we are ready to see how Hume applies his test for meaning to “substance” and to “self.” In order to keep our discussion anchored to Hume's texts, we shall use his own formulation of the test: from what impression(s) are the ideas of substance and of self derived? It should always be remembered, however, that this question could be reformulated in terms of modern meaning-empiricism: can the terms “substance” and “self” be defined ostensively, by indicating something in experience that would answer to them?

In the following paragraph from a section of the *Treatise* (T:16) entitled “Of Modes and Substances,” Hume applies his test for meaning to the notion of substance:

I wou'd fain ask those philosophers, who found so much of their reasonings on the distinction of substance and accident, and imagine we have clear ideas of each, whether the idea of *substance* be deriv'd from the impressions of sensation or reflexion? If it be convey'd to us by our senses, I ask, which of them; and after what manner? If it be perceiv'd by the eyes, it must be a colour; if by the ears, a sound; if by the palate, a taste; and so of the other senses. But I believe none will assert, that substance is either a colour, or sound, or taste. The idea of substance must therefore be derived from an impression of reflexion, if it really exist. But the impressions of reflexion resolve themselves into our passions and emotions; none of which can possibly represent a substance. We have therefore no idea of substance, distinct from that of a collection of particular qualities, nor have we any other meaning when we talk or reason concerning it.

Hume's point, of course, is that “substance” fails to pass his test for meaning. For in order for this term to have meaning for us, we must have an idea of substance. Now, argues Hume, if we have an idea of substance, then that idea must be derived either from an impression of sensation or an impression of reflection. So, using the strategy that we shall see him use again for “self” and for “cause,” Hume challenges philosophers who rely on the notion of substance

to specify the impression from which it is derived. If it is derived from an impression of sensation, then substance must be something we perceive through our senses, for instance, as a color or a sound or a taste. But substance cannot possibly be a color or a sound or a taste, since these are supposed to be qualities *of* substances. So, the idea of substance is not derived from any impression of sensation. If, on the other hand, it is derived from an impression of reflection, then substance must be something we are aware of by introspection – some conscious state such as an emotion or a feeling. But no such introspectible conscious states can possibly be equated with a substance, since they are supposed to be properties of a (mental) substance. Therefore, Hume concludes, we have no idea of substance, and the term “substance” is meaningless. In the final sentence of the passage, Hume concludes that we have no choice but to opt for the bundle theory: a thing is just a bundle of properties (a “collection of particular qualities”).

Put in terms of modern meaning-empiricism, Hume’s argument would go as follows. We cannot find anything in our experience, whether through sense perception or introspection, that answers to the term “substance.” For substance is supposed to be something that has, supports, or underlies all of a thing’s observable properties, but that is not itself perceivable. If it were perceivable, it would not be a substance but only a property. Substance itself is unperceivable in principle. Therefore, the term “substance” fails to pass the modernized Empiricist test for meaning and is, accordingly, meaningless.

As we have said, seventeenth- and eighteenth-century philosophers commonly held, following Descartes, that there is such a thing as a purely mental substance – one that has no physical properties like shape or size, but whose only properties are its conscious states. It might perhaps be thought, then, that even if Hume has shown that we have no idea of material or physical substance, we still have an idea of mental substance. Indeed, one major philosopher who strenuously rejected the notion of material substance, Bishop Berkeley, nevertheless retained and even glorified the notion of mental substance.⁷ But Hume, with ruthless consistency, applies his test for meaning to this notion as well. He does this in a section of the *Treatise* called “Of personal identity.” In that section, Hume applies his test to the notion of *self*. But the notion of self that he has in mind is obviously that of a mental substance, mind, or soul; so that he is in effect applying his test to the notion of mental substance used by Descartes and most other seventeenth- and eighteenth-century philosophers. Here is part of what Hume says:

There are some philosophers, who imagine we are every moment intimately conscious of what we call our SELF; that we feel its existence and its continuance in existence; and are certain, beyond the evidence of a demonstration, both of its perfect identity and simplicity

Unluckily all these positive assertions are contrary to that very experience, which is pleaded for them, nor have we any idea of self, after the manner it is here explain’d. For, from what impression cou’d this idea be deriv’d? This question ’tis

impossible to answer without a manifest contradiction and absurdity; and yet 'tis a question, which must necessarily be answer'd, if we wou'd have the idea of self pass for clear and intelligible. It must be some one impression that gives rise to every real idea. But self or person is not any one impression, but that to which our several impressions and ideas are suppos'd to have a reference. If any impression gives rise to the idea of self, that impression must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner. But there is no impression constant and invariable. Pain and pleasure, grief and joy, passions and sensations succeed each other, and never all exist at the same time. It cannot, therefore, be from any of these impressions, or from any other, that the idea of self is deriv'd; and consequently there is no such idea

I may venture to affirm . . . that [a mind is] nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in perpetual flux and movement. Our eyes cannot turn in their sockets without varying our perceptions. Our thought is still more variable than our sight The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations The comparison of the theatre must not mislead us. They are the perceptions only, that constitute the mind; nor have we the most distant notion of the place, where these scenes are represented, or of the materials, of which it is compos'd.

(T:251–3)

In the second paragraph of the passage just quoted, Hume makes two points. One is that if one tries to pinpoint the impression from which the idea of one's own mind or self is derived, one simply does not find it. All one finds, instead, is a constantly shifting vista of sense impressions, feelings, images, etc. One does not find, in addition to these, any single enduring unchanging impression that one could identify as one's own self or mind. In other words, if one tries to spot one's own self or mind by introspection, one just cannot find it. As Hume puts it (T:252):

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and can never observe anything but the perception If any one, upon serious and unprejudic'd reflexion, thinks he has a different notion of *himself*, I must confess I can reason no longer with him.

Hume's other point is that it is not even *possible* for there to be an impression of self. For "self or person is not any one impression, but that to which our several impressions and

ideas are supposed to have a reference.” What Hume is saying can be put this way. To try to spot one’s self by introspection is to try to introspect the very subject of consciousness – the subject which is conscious of all the objects that one is aware of. But this is like trying to see the point *from which* one sees everything. The attempt is bound to fail, for that point cannot be seen; it is the one point which can never be in one’s visual field. Thus, not only is there no impression of the self; there cannot be one. It follows from each of these two points, together with Hume’s test for meaning, that the notion of a self, understood as a mental substance or mind that has various conscious states as its properties, is meaningless. It is just as empty as is the notion of a material substance that has various qualities as its properties.

In the last of the paragraphs quoted above, Hume draws his conclusion. The only tenable view of the mind or self is in terms of a bundle theory: the mind is “a bundle or collection of different perceptions They are the successive perceptions only, that constitute the mind.” A mind, then, is nothing but a succession of conscious states.

When Hume’s test for meaning is applied to the notion of a material substance, it yields the result that the notion is meaningless and that a thing can be only a bundle of properties. Likewise, when the test is applied to the notion of a mental substance, it yields the result that this notion is meaningless and that a mind can be only a bundle of conscious states. Meaning-empiricism leaves no room at all for the notion of substance as distinct from its properties, whether it be a material substance or a mind. In the next two parts, we shall consider whether these results are philosophically acceptable.

3 An alternative to the substance theory

Meaning-empiricism implies that the substance theory should be rejected. We have seen, however, that the substance theory rests on an argument, namely the argument from change. If we wish to reject the substance theory, therefore, we must be prepared to refute that argument. One cannot both accept an argument for a theory and reject that very theory. The time has come, therefore, for us to evaluate the argument from change. The argument, you will recall, goes like this:

- (1) We can distinguish between (a) all of a thing’s determinate properties changing without the thing’s ceasing to exist, and (b) a thing’s ceasing to exist.
- (2) We can distinguish between (a) and (b) only if a thing is composed, in addition to its properties, of a permanent underlying substance.

∴ A thing is composed, in addition to its properties, of a permanent underlying substance.

Premiss (1) might be challenged, in the following way. Can we really distinguish between *all* of a thing’s determinate properties changing without the thing ceasing to exist, and the

thing ceasing to exist? In order for a thing to continue existing, must it not retain at least some of its determinate properties – if only a single one? And isn't this condition, in fact, satisfied in the case of Descartes' piece of wax? For although all of the wax's observable properties have changed, does not the wax still retain a determinate chemical composition – or perhaps a determinate atomic structure – by virtue of which it continues to exist despite the change in the properties that can be observed with the unaided senses?

Although this line of attack certainly has some force, and would be endorsed by those contemporary philosophers who call themselves "scientific realists," there are other contemporary philosophers who would oppose it. For suppose, as is quite conceivable, chemists or atomic physicists discovered that when wax is melted, its color, shape, consistency, and other observable properties are not the only ones that change. Rather, its chemical composition and/or atomic structure are *also* altered (perhaps only very slightly so). Would we then have to agree that the piece of wax Descartes describes ceases to exist when it is melted? Opponents of scientific realism would answer that we would not – at least not merely on the strength of what we have supposed so far, namely, that both the wax's more obvious properties and its "scientific" properties have altered. They would hold that if, by cooling and moulding the melted stuff, we could easily get it back to its previous shape, size, texture, and consistency, and do such things as use it for a candle or as a water repellent, then it would still be wax, even if its atomic or chemical structure had altered. If this is right, then what accounts for the wax continuing to exist when melted is not that it retains "scientific" properties like its chemical composition or its atomic structure; those could change, as well as the more obvious properties of shape, size, hardness, color, odor, and so on. So, the opponent of scientific realism would say, premiss (1) withstands the challenge that was raised against it.

A committed scientific realist, however, would probably not be impressed by this line of thought, but would insist that a thing simply cannot continue to be wax if it loses the atomic or chemical structure of wax. Rather than dismissing this view as false, accordingly, we shall maintain that if scientific realism is correct, then premiss (1) of the argument from change need only be qualified, along the following lines:

- (1a) We can distinguish between (a) all of a thing's determinate properties *except* for scientific properties P changing without the thing's ceasing to exist, and (b) a thing's ceasing to exist.

Here "scientific properties P" denotes whatever determinate scientific property or properties the scientific realist holds that a thing must retain in order to continue existing. As I will argue, however, there is good reason to think that even if (1) is false, then at least this qualified version of (1) is true.

What I have to say about (1a), however, will be clearer if first we critically examine the *second* premiss of the argument from change, to which we therefore now turn our attention.

One might challenge the “only if” in premiss (2), that is, the very strong claim that we can distinguish between (a) all of a thing’s determinate properties changing without the thing’s ceasing to exist and (b) the thing’s ceasing to exist, in *only one* way – i.e. *only if* the substance theory is true. The best way to make good such a challenge, of course, would be to provide an alternative account of how this distinction can be made – an account that does not involve substance. For premiss (2) operates as a kind of challenge: how can we justify saying that a thing, all of whose determinate properties have changed, has not itself ceased to exist *unless* we postulate an underlying substance that has not changed? Clearly if an Empiricist can give an alternative account of the distinction between (a) and (b), then this challenge will have been met.

As a preliminary to offering such an account, we will raise an objection to premiss (2). This is that “substance” is actually quite incapable of explaining our ability to distinguish (a) from (b). For how do we actually make this distinction? The answer, surely, is that we make it on the basis of the changes that we observe. Speaking very roughly, if the changes we observe in a thing are not too radical, then we say that it has changed its properties without losing its identity; whereas if the changes we observe are more radical, then we say that the thing has lost its identity or, which amounts to the same thing, that it has ceased to exist. Notice, then, that when we actually determine whether a thing has merely undergone a change of properties or has ceased to exist, “substance” plays no role: it is simply irrelevant. Furthermore, “substance” *could not* help; it is *useless*. For how could an unperceivable substance possibly help us to make a distinction that we make on the basis of what we observe? It would seem, then, that we have already found good reason to doubt premiss (2) of the argument from change.

In order to give an alternative account of how (a) is distinguished from (b), an Empiricist needs to offer an account of what it is for a thing to retain its identity while undergoing change, or of “identity through change.” In other words, the Empiricist needs to answer the following question: what is required for a thing to continue to be the same thing, during a period of time in which it changes? For if one can specify such a requirement, then one can hold that whether we have a case of (a) or of (b) turns on whether that requirement is satisfied.

Hume himself had a response to this question, but we shall not try to defend it. For basically, Hume’s response was to reject the question, on the grounds that in order for a thing to retain its identity, it *cannot change at all*. So, for example, if a few planks in a ship’s hull were replaced, it would no longer be the same ship; or if one new bolt were added to your car, it would no longer be the same car. Hume bases this surprising view on an exceedingly narrow definition of identity through time, according to which such identity means that a thing is not only “uninterrupted,” but also “invariable” or unchanging (T:235; see also T:201–2). He says therefore that when we ascribe identity to something that has changed, we are simply mistaken. He adds, however, that there are certain features of things that induce us to make this mistake: namely, their changing only in small ways or in proportionally small ways, or

gradually, or their having parts that serve a unified function (a “common end or purpose”), or parts that are interdependent and serve such a function, or their having temporal stages that exhibit qualitative sameness or “specific identity” (as opposed to numerical identity), or their being by nature changeable. These features, he says, persuade us to mistake what is really a succession of related things for one and the same thing, because the act of mind by which we consider such a succession of related things closely resembles or feels very like the act of mind by which we consider an uninterrupted and unchanging thing (T:6, 255–8). Some recent philosophers would say that the very features Hume mentions are in fact those that *justify* us in ascribing to things identity through change, rather than, as he thought, features that *mislead* us into ascribing identity to things that are really diverse. Seen in this way, Hume's discussion is illuminating, because the features he insightfully locates are indisputably relevant to our ordinary judgments of identity.⁸

Despite Hume's insightfulness, however, his narrow definition of identity prevents him from taking seriously the question of what is required for a thing to retain its identity even though it has changed. Rather than expounding Hume's view in more detail, therefore, we shall sketch out an answer, suggested by Locke (1975) in Chapter 27 (titled “Of Identity and Diversity”) of Book II of his *Essay Concerning Human Understanding*, and favored by many empirically minded contemporary philosophers. This is the view that the fundamental requirement for identity through change is *spatiotemporal continuity*.

The requirement of spatiotemporal continuity is best introduced negatively, by considering a case in which it would be violated. Here, then, is such a case. Suppose that on a table at the front end of a room, there is a large, heavy object – say a marble statue of George Washington. Next, suppose that this statue ceases to occupy its place on the table and that, some time afterwards, a statue exactly like it begins to occupy a place at the back end of the room. Furthermore, imagine (this requires indulging in a bit of science fiction) that the statue did not occupy a series of contiguous places between the front of the room and the back of the room; nor did it somehow get transformed into energy and occupy such a series of places in that form, before rematerializing as a marble statue when it reached the back of the room. Rather, what happened is simply that the statue ceased to occupy its place at the front of the room, and later a statue exactly like it began to occupy a place at the back of the room. Now consider this question: was the statue at the front of the room the same statue as the one at back of the room? Here it is important to understand what this question is asking. It is not asking whether the statue at the front of the room was *qualitatively* the same as the one at the back. By hypothesis, it was the same: we just said that the statue in the back was “exactly like” the one in the front; meaning that it had the same weight, color, shape, size, workmanship, etc. So, it was qualitatively the same – it had exactly the same qualities or properties – as the one at the front. The question is, rather, whether the statue at the front of the room and the one at the back of the room were *numerically* the same: was there just one statue which existed at the earlier time and also at the later time, or were there two statues, one at the earlier time and another at the later time? (Was the situation like that of one person at two

different times, or like that of identical twins? Was the statue at the front of the room the *same one* as the statue at the back, or only the *same as* the one at the back?) It should be clear, perhaps after a bit of reflection, that the answer is that the statue at the back of the room and the one at the front of the room were not numerically identical. They were qualitatively identical, but numerically distinct (like the twins). The statue at the front of the room was the same as the one at the back, but it was not the same *one* as the one at the back. The reason is that between the time the statue at the front of the room ceased to occupy its place and the one at the back began to occupy its place, there was an interruption or break in spatiotemporal continuity. The statue at the front of the room was not spatiotemporally continuous with the one at the back, as it would have been if, for example, someone had carried it from one place to the other.

At this point, you may feel like raising the following objection. “The case of the ‘vanishing statue’ just presented is a science-fiction case; it couldn’t really happen. So, how can it be relevant to our question?” The answer is that in considering conceptual questions, such as the question of what is required for a thing to retain its identity through time, unrealistic cases are sometimes indispensable. To see this, let us return to Descartes’ case of the melting wax. In most cases in which a thing’s properties change without the thing losing its identity, what happens is that only a few of its determinate properties change. For example, a leaf changes color, or develops a hole, or falls from a tree. In such cases, we have no trouble determining that the leaf is still the same leaf: it just hasn’t changed enough for the question even to arise of whether it has ceased to exist. On the other hand, in cases where a thing changes so much that it does lose its identity, what generally happens is that the properties change quite drastically. For example, a piece of paper is burned until it is just a pile of ashes. In that case, we have no trouble determining that the paper has been destroyed: it has changed in ways that prevent it from continuing to exist as a piece of paper. Descartes’ case of the wax, however, sits uneasily between those two clear cases. On the one hand, it is like the burning paper example in that the wax’s determinate properties have changed drastically. On the other hand, it is like the leaf case in that the wax does not get destroyed.

So, in order to get clear on what is required for a thing to retain its identity through change, we can ask: what more would have to happen in order for the wax to be destroyed? Well, we know that if it changed in still other ways – if it underwent certain chemical reactions that broke it down into separate elements – it would be destroyed. But is there anything else that could conceivably happen, besides such a chemical change or some other dramatic change in the wax’s properties, that would make the wax cease to exist? If we can specify something else, then we will have unearthed a condition of identity through change other than any condition pertaining to the loss and gain of properties. Well then: suppose that the wax had ceased to occupy its place, and that a puddle of melted wax, or even a piece of unmelted wax exactly like it, had begun to occupy a different place, in the same way as our “vanishing statue.” *Of course* this doesn’t really happen: we are in the realm of science fiction again. But suppose it did happen. It is, after all, perfectly conceivable; it is a logical possibility. Then,

would we not admit that the original piece of wax had ceased to exist, and that another piece of wax had begun to exist? It seems clear enough that we would. But then this shows that spatiotemporal continuity is one of our requirements for identity through change.

Indeed, this requirement appears to be fundamental. For suppose that scientific realists are right, and that a thing cannot retain its identity if it loses certain determinate "scientific" properties, like its atomic or chemical structure. In that case, as we saw, premiss (1) of the argument from change would have to be reformulated to say:

- (1a) We can distinguish between (a) all of a thing's determinate properties, *except* for scientific properties P, changing without the thing's ceasing to exist, and (b) a thing's ceasing to exist.

Is (1a) true or false? The answer seems to be that (1a) is true; for if we suppose that a piece of wax ceased to occupy a certain place, p_1 , and that one exactly like it later began to occupy another place, p_2 (in the same way as our "vanishing statue"), *and that the wax in both places had the chemical or atomic structure of wax*, it would still be true that the piece of wax that had occupied p_1 was numerically different from the one that occupied p_2 : we would then have a case of (b) rather than a case only of (a), just because of the break in spatiotemporal continuity. Thus even if scientific realism is right, it does not follow that spatiotemporal continuity is not a necessary condition for identity through time. Rather, what follows is that keeping the same molecular or atomic structure is *also* a necessary condition of identity through time.

Having introduced the requirement of spatiotemporal continuity, let us formulate it explicitly, and give a definition of spatiotemporal continuity itself. We can formulate the requirement as follows:

X at time t_1 is the same thing as Y at time t_2 only if X is spatiotemporally continuous with Y from t_1 to t_2 .

Here both "X" and "Y" denote one specific object of a specific kind, such as the book you are holding now and the book that is in your knapsack later. In cases where both "X" and "Y" designate the very same thing considered at one moment of its existence, t_1 and t_2 will refer to the same time, and the condition of spatiotemporal continuity will be automatically satisfied in virtue of the fact that X and Y also occupy exactly the same place at that time. In cases like those which we are chiefly concerned with, where X is the same thing as Y despite having undergone change, t_1 and t_2 must of course refer to different times. X could be Descartes' unmelted lump of wax and Y the melted wax. The requirement then states that in order for the unmelted wax to be the same thing as the melted wax, the unmelted wax must be spatiotemporally continuous with the melted wax.

The definition of spatiotemporal continuity itself is more difficult to give. The basic notion is that of uninterrupted existence in space and time – a condition that was *not* met in our case of the “vanishing statue.” But, while it seems easy enough to have an intuitive understanding of this notion, it is difficult to say precisely what the notion comes to. We shall now try to give an exact definition, but since this definition will have to be somewhat technical, the following four paragraphs may be skipped (or returned to later) without risk of losing the main thread of argument being developed.

Let us try the following definition where, again, “X” and “Y” stand for particular objects of specific kinds, but where “an X” or “a Y” stand for any objects of specific kinds:

DI: X is spatiotemporally continuous with Y from t_1 to $t_2 = df.$

- (1) X occupies p_1 at t_1 ;
- (2) Y occupies p_2 at t_2 ;
- (3) either (i) p_1 is identical with p_2 and, from t_1 to t_2 , either an X or a Y occupies p_2 ; or (ii) p_1 is not identical with p_2 , but there is a series of contiguous places, $S_1 \dots S_n$, such that (a) S_1 is contiguous with p_1 and S_n is contiguous with p_2 , (b) at every time from t_1 to t_2 every contiguous member of $S_1 \dots S_n$ is successively occupied by either an X or a Y.

Clause (3) is intended to cover two different possible situations: (i) that in which a thing changes while remaining stationary; and (ii) that in which a thing both changes and moves. Suppose for example that X is a particular lump of unmelted wax and Y is a particular lump of partially melted wax. Then there are two ways in which clause (3) can be satisfied. *Either* the place occupied by the unmelted wax is the same as the place occupied by the partially melted wax, and from t_1 to t_2 this place is always occupied by either unmelted wax or partially melted wax – in which case we have a stationary piece of melting (or perhaps alternately melting and solidifying) wax. *Or* there is a series of contiguous places between the place occupied by the unmelted wax and the place occupied by the partially melted wax, such that:

- (a) the first member of the series is contiguous with the place occupied by the unmelted wax and the last member is contiguous with the place occupied by the partially melted wax;
- (b) at all times from t_1 to t_2 , every contiguous member of this series of places is successively occupied by either unmelted wax or partially melted wax – in which case we have a moving piece of melting (or perhaps alternately melting and solidifying) wax.

DI, however, will not quite do. For consider a case where a piece of wax gradually melts until it is completely liquid. We want the definition to let us say that the unmelted wax is

spatiotemporally continuous with the completely melted wax (for otherwise, we will not be able to say, given that spatiotemporal continuity is a requirement for identity through change, that it is the same piece of wax). But it does not. For in such a case, clause (3) will not be satisfied, whether the wax remains stationary or moves. If it remains stationary, then there will be a time between t_1 and t_n when the place the wax occupies is occupied by neither unmelted nor totally melted wax, but by partially melted wax instead, so (3i) will not be satisfied. (Nor, of course, will (3ii) be satisfied, since it requires that X be at a different place at t_1 from Y at t_2 , which cannot happen if the wax remains stationary.) If the wax moves, then between t_1 and t_2 there will be at least two contiguous members of $S_1 \dots S_n$, such that they are not successively occupied by either unmelted or totally melted wax, but rather by unmelted wax and partially melted wax, or by (totally) melted wax and partially melted wax; so (3ii) will not be satisfied. (Nor, of course, will (3i) be satisfied, since it requires that X at t_1 be at the same place as Y at t_2 , which cannot happen if the wax moves.)

To get around this difficulty, we must expand the definition as follows:

D2: X is spatiotemporally continuous with Y from t_1 to $t_2 = df.$

Either:

- (1) X occupies p_1 at t_1 ;
- (2) Y occupies p_2 at t_2 ;
- (3) either (i) p_1 is identical with p_2 and, from t_1 to t_2 , either an X or a Y occupies p_2 ; or (ii) p_1 is not identical with p_2 , but there is a series of contiguous places, $S_1 \dots S_n$, such that (a) S_1 is contiguous with p_1 and S_n is contiguous with p_2 , (b) at every time from t_1 to t_2 every contiguous member of $S_1 \dots S_n$ is successively occupied by either an X or a Y.

Or

- (4) X is spatiotemporally continuous with Z_1 from t_1 to t^* , and Z_1 is spatiotemporally continuous with Y from t^* to t_2 ; or X is spatiotemporally continuous with Z_1 from t_1 to t^* , Z_1 is spatiotemporally continuous with Z_2 from t^* to t^{**} , and Z_2 is spatiotemporally continuous with Y from t^{**} to t_2 ; or X is spatiotemporally continuous with Z_1 from t_1 to t^* , Z_1 is spatiotemporally continuous with Z_2 from t^* to t^{**} , Z_2 is spatiotemporally continuous with Z_3 from t^{**} to t^{***} , and Z_3 is spatiotemporally continuous with Y from t^{***} to t_2 ; or . . . and so on.

More simply, the last clause says that X must be spatiotemporally continuous with something which is itself spatiotemporally continuous with Y, or with a member of a series each of whose members is spatiotemporally continuous with the next member, and whose last member is spatiotemporally continuous with Y. This covers the case of the gradually melting wax, because the unmelted wax is spatiotemporally continuous with the partially melted

wax, which is spatiotemporally continuous with wax in a still more advanced state of melting, and so on. In other words, the totally unmelted wax is spatiotemporally continuous with a member of a series whose members are in progressively more advanced stages of melting, each of which is spatiotemporally continuous with the next, and the last of which is spatiotemporally continuous with the totally melted wax.

D2 is an example of a kind of definition called a *recursive* definition. For its *definiendum* (the term being defined) recurs within the *definiens* (the terms doing the defining). Thus the term, “— is spatiotemporally continuous with — from — to —,” which is the one being defined, recurs in (4). The definition is not circular, however, because each occurrence of this term in (4) is completely defined in terms of the definiens’s first disjunct, that is, in terms of clauses (1), (2), and (3).

The requirement of spatiotemporal continuity provides at least a partial answer to the challenge posed by the argument from change. For if that requirement is *not* satisfied, then we have a case of a thing ceasing to exist, rather than a case of a thing’s properties changing. (Actually, for this claim to be true in all cases, the requirement must be slightly amended. The reason why, and the amendment needed, will be given at the very end of this part.) Notice also that the requirement of spatiotemporal continuity is an *empirical* requirement: we can determine by observation whether or not it is satisfied.

But can we also say that whenever the requirement is satisfied, then we have only a case of a thing’s properties changing? No. For consider again the case of the paper that burns down to a heap of ashes. In that case, the requirement is satisfied: the piece of paper is spatiotemporally continuous with the heap of ashes. Yet, the paper has certainly lost its identity, or been destroyed. It would be easy to describe other similar cases. Such cases are important, for they show that although, if X is the same thing as Y, X is spatiotemporally continuous with Y, it is false that, if X is spatiotemporally continuous with Y, X is the same thing as Y. In other words, although X is the same thing as Y *only if* X is spatiotemporally continuous with Y, we cannot correctly say that X is the same thing as Y *if* (just because) X is spatiotemporally continuous with Y. Nor, therefore, can we say that X is the same thing as Y *if and only if* X is spatiotemporally continuous with Y. To put this point still another way: spatiotemporal continuity is a *necessary condition* for identity through change, but it is not also a *sufficient condition* for identity through change.

Then what conditions *are* sufficient? What, in addition to spatiotemporal continuity, would ensure that X is the same thing as Y? Our answer to this question – an answer that derives again from Locke’s account of identity and diversity – may be a bit disappointing. It is that the conditions that are sufficient depend on the *sort* of thing we are talking about. For example, suppose that we are talking about a molecule of water. Then not only must it satisfy the continuity requirement, but it must retain exactly two atoms of hydrogen and one atom of oxygen; take away either the atom of oxygen or one of the atoms of hydrogen, and you have destroyed the molecule of water. On the other hand, suppose we are talking about a tree. Then not only must it satisfy the continuity requirement, but it must retain a certain configuration involving trunk, branches, etc. Except for spatiotemporal continuity, there is

so to speak no exact “common measure” between the two cases. Again, it would be easy to give further examples illustrating the same point, and this shows that no specific sufficient condition of identity through change can be given that applies generally or across the board. What one can do, however, is to give a necessary condition (spatiotemporal continuity), and to indicate by examples how the conditions that are sufficient vary from one sort of object to another. From these examples, we can also extract certain features that are often relevant to a thing maintaining its identity through change, such as the features Hume identified: if the thing changes slowly, or in small ways, or in ways that preserve its function, etc. But we cannot say that any set of these features is sufficient in all cases; we cannot give a “recipe” that tells us what, beyond spatiotemporal continuity, always ensures that a thing preserves its identity.

If what we have said is correct, then there are three noteworthy consequences. First, the argument from change may now be rejected as unsound, since its second premiss is false. So, we are free to accept a bundle theory, on which a thing is a collection of property instances exhibiting spatiotemporal continuity, plus whatever other conditions are appropriate to the sort of thing that it is. Second, the substance theory commits a double error. For it purports to give a condition – retaining the same underlying substance – which is *both necessary and sufficient* for retaining an identity through change. But the condition is not necessary, and the attempt to give a sufficient condition is misconceived. Third, our discussion suggests that although we cannot give a condition for identity over time which is both necessary and sufficient, we can give at least the general form of such a condition; provided that we frame the question, not as “what is the general form of the necessary and sufficient conditions for X at t_1 to be the same thing as Y at t_2 ?”, but rather as “what is the general form of the necessary and sufficient conditions for X at t_1 to be the same F as Y at t_2 ?” where F designates a kind or sort of thing. (For example, instead of asking for the necessary and sufficient conditions for the unmelted wax to be the same *thing* as the melted wax, we ask for the necessary and sufficient conditions for the unmelted wax to be the same *wax* as the melted wax.)

We can then answer our question as follows: X at t_1 is the same F as Y at t_2 if and only if:

Either

- (1) X is spatiotemporally continuous with Y from t_1 to t_2 ; and
- (2) X is a thing of kind *F* at t_1 and Y is a thing of kind *F* at t_2 ; and
- (3a) for all times t_x and t_y between t_1 and t_2 , there is no Z such that Z is spatiotemporally continuous with X or with Y from t_x to t_y and Z is not a thing of kind *F*.

Or

- (3b) If X and Y are composed of parts, then a set of parts of X and of Y – such that, if those parts were assembled in a certain way, they would then constitute a thing of kind *F* – is such that each of its members satisfies conditions (1), (2), and (3a).

Clause (3b) is needed to cover cases of the sort: my watch is disassembled, the parts are sent for repair to different locations, and the watch is reassembled. Without clause (3b), we could not say that the reassembled watch is the same watch as the original one. But (3b) allows us to say that so long as the watch is composed of parts each of which individually meets conditions (1), (2), and (3a) – as would, for example, one of the watch’s gears, just by virtue of exhibiting spatiotemporal continuity and being a *gear* from the time of disassembly to the time of reassembly – the reassembled watch is the same watch as the original. Finally, notice that such cases, involving disassembly, also show that the basic requirement of spatiotemporal continuity must be amended, as follows:

X at time t_1 is the same thing as Y at time t_2 only if either X is spatiotemporally continuous with Y from t_1 to t_2 or some part(s) of X are spatiotemporally continuous with some part(s) of Y from t_1 to t_2 .

4 Critique of Hume’s bundle theory of the self

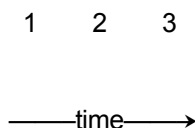
The foregoing discussion has shown that the notion of substance, at least with respect to material things, need not be accepted. For it is possible to give an account of such things’ identity through change without postulating an underlying permanent substance. So, the bundle theory (now elaborated as saying that a thing is a collection of instantiated properties exhibiting spatiotemporal continuity plus whatever other features are appropriate to the kind of thing it is) seems to be tenable and, given the objections to the substance theory that have been discussed, preferable to the substance theory. As we have seen, however, Hume also offers a bundle theory of the self or mind. Is such a theory tenable? Our thesis in this part will be that it is not.

One way to approach the issue would be to take a course parallel to that taken for material things. We could proceed by asking: what, other than mental substance, is required for a person to retain an identity over time and through change? This question, which is called “the problem of personal identity,” has been discussed by both classical and contemporary philosophers. Hume, operating with his narrow definition of identity, held that personal identity is nothing but a fiction. Locke, in a classic and very influential treatment of the issue, tried to define personal identity purely in terms of memory (see Locke 1975: Book 2, Chap. 27). Today, some philosophers believe that a person retains personal identity as long as he or she retains the same body – or at least the same brain. Perhaps most contemporary philosophers think that both memory and the body must enter into an adequate account of personal identity.

Although the problem of personal identity is a fascinating one, we shall not delve into it, because to do so would require digressing at too great a length from our study of Hume.⁹ Instead, we shall limit ourselves to advancing a single objection to Hume’s bundle theory of the mind. In fairness to Hume, it should be noted that he may have been aware of the

objection. For in his “Appendix” to the *Treatise* (T:633–6), he himself expressed dissatisfaction with his bundle theory of the self, and even retracted the theory. But he also declared himself unable to offer a better theory; for, as he acknowledged, the theory is the logical outcome of applying his test of meaning to the notion of mental substance – a test that Hume remained unable or unwilling to give up. Despite Hume’s disavowal of his own bundle theory of the self, therefore, we need to see for ourselves what is wrong with the theory, and what the implications are for meaning-empiricism.

Our objection to the bundle theory may be called “the argument from the awareness of succession in time.” Historically, it derives from Immanuel Kant’s *Critique of Pure Reason* (1781), but this treatment is more directly inspired by the lucid presentation given by the twentieth-century English philosopher Charles A. Campbell (1962: 224–235). Suppose, to use Campbell’s example, that you hear three successive strokes of a church bell.



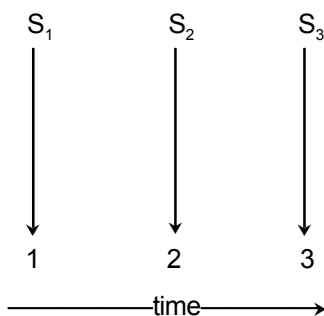
Clearly, this is something that all of us with normal hearing can do: we can all hear the first stroke, and then hear the second stroke *as coming after the first*, and then the third stroke as *coming after the first two*. Indeed, this case is just a simple example of a basic fact about human awareness or consciousness: it is always successive or durational, and not just a momentary affair, like a spark. Admittedly, this fact does not appear to be a necessary feature of any consciousness whatsoever. A state of perpetually renewed amnesia is not logically impossible: there could be a person who at every moment forgot the preceding moment. Such a person would never have the experience of succession in time. He or she could not hear the strokes of the church bell as a succession – could not hear the second stroke as coming after the first, or the third as coming after the first and the second. For by the time this person heard the second stroke, the first stroke would have been forgotten, and by the time the person heard the third stroke, the second would have been forgotten. Nor could such a person hear a melody, or understand a sentence; he or she could have only the most primitive, momentary awareness – say a sensation of pain or of pleasure – which would be promptly forgotten forever. Such an impoverished awareness is no doubt possible; but happily it isn’t the sort of awareness that we in fact have. Awareness is, for us, awareness of one state of things coming after another. Any theory of the self or the mind that is incompatible with this basic fact is untenable.

Is Hume’s theory compatible with awareness of succession? According to the bundle theory, a mind or self is nothing but a series of perceptions. In terms of our example, these perceptions include each of the auditory perceptions of the bell, plus all the previous perceptions beginning with those had in infancy, and all the subsequent ones up until death.

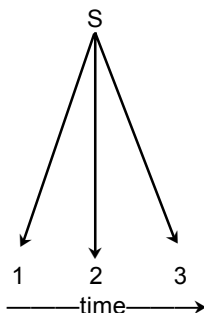
But there is no enduring mind or self, distinct from these perceptions. Does this allow for awareness of succession in time? Could a self or mind, as described by Hume, experience the second stroke as the successor of the first, and the third as the successor of the second?

To begin with, it is questionable whether Hume's theory even allows for awareness of each stroke *singly*. Recall that Hume holds that there cannot be any impression of the conscious subject, any more than there can be a visual impression of the point from which one sees. But if there cannot be an impression of the conscious subject, then there can be no idea of it either; so the notion ought to be rejected as meaningless. So Hume's principles commit him not only to a bundle-theory of the mind but, so to speak, to a "no-subject bundle theory." But if there is no conscious subject, then *who* or *what* is aware of each stroke of the bell? The answer, it would seem, is "no one." But then there is no awareness of the strokes, even singly; for awareness of X is a two-term relation which requires not just X, but also a conscious subject who is aware of X.

Suppose, however, that we amend Hume's theory, by supplying a conscious subject who is aware of each stroke of the bell. In other words, suppose that we change Hume's theory into a "subject bundle theory", or better put, a "bundle-of-subjects" theory, represented by the following diagram:



This allows for awareness of each stroke singly. But does it allow for awareness of them as successive? It does not. For if the strokes are to be heard as successive, then all of them must be heard by the same conscious subject, who must continue to exist at least from the beginning to the end of the succession. The situation must be as represented by this diagram:



But on the “bundle-of-subjects” theory, the strokes are not heard by the same conscious subject; rather, each stroke is heard by a different conscious subject. Therefore, they are not heard as coming one after another. It is much as if our victim of perpetually renewed amnesia had been listening to the bell. The argument from the awareness of succession in time, then, is this. Awareness of succession in time requires a conscious subject who continues to exist at least from the beginning to the end of the succession; but, on Hume’s theory, there is no such subject. Therefore, Hume’s theory does not allow for awareness of succession in time.

As mentioned earlier, this argument derives from Kant’s *Critique of Pure Reason* (in a famous and very difficult section called “The Transcendental Deduction of the Pure Concepts of the Understanding”). Later in the same work (in a section called the “Paralogisms of Pure Reason”), Kant cautions against drawing illegitimate conclusions from the argument. He points out, for example, that it does not follow that the conscious subject is an immaterial substance, or that it is indestructible, immortal, or anything of the sort. More basically, Kant points out that from this argument alone we cannot tell anything about the *nature* of the conscious subject. We can only know there must be such a subject, since that is a necessary condition of any awareness of succession in time. But our notion of this subject (to which Kant gives the imposing name “the transcendental unity of apperception”) is devoid of any specific content: it is simply that which is successively aware of various items.

In somewhat the same spirit as Kant’s, we should not draw illegitimate conclusions from the failure of Hume’s bundle theory of the mind. In particular, we should not conclude that the Empiricist view of meaning that led to this bundle theory is completely mistaken; for an Empiricist could modify the view as follows. Instead of requiring that every meaningful descriptive term be either ostensively definable, or definable by means of terms that are themselves ostensively definable, the Empiricist could broaden the view to allow also that any term designating something which is a logically necessary condition of the sorts of experience that we have, is also meaningful. This would allow the term “conscious subject” and its cognates to be meaningful, without allowing the term “substance underlying a thing’s observable properties” and its cognates to be meaningful. It would preserve the linkage of meaning to experience upon which meaning-empiricism insists.

HUME'S THEORY OF KNOWLEDGE (I)

“Hume’s Fork”

1 Introduction

This chapter will discuss Hume’s division of knowable propositions into two classes: “relations of ideas” and “matters of fact.”¹ This division, which has come to be known as “Hume’s Fork,” is fundamental both to Hume’s empiricism and to the twentieth-century empiricism that it inspired.² Hume presents it in the first two paragraphs of Section IV of the *Enquiry* (hereafter designated simply as “Section IV”). It sets the stage for the seminal critique of causal and inductive reasoning that occupies the rest of Section IV, which will be discussed in Chapter 3.

2 “Relations of ideas” and “matters of fact”

The first two paragraphs of Section IV of the *Enquiry* are as follows:

All the objects of human reason or enquiry may naturally be divided into two kinds, to wit, *Relations of Ideas*, and *Matters of Fact*. Of the first kind are the sciences of Geometry, Algebra, and Arithmetic; and, in short, every affirmation, which is either intuitively or demonstratively certain. *That the square of the hypotenuse is equal to the squares of the two sides*, is a proposition which expresses a relation between these figures. *That three times five is equal to the half of thirty*, expresses a relation between these numbers. Propositions of this kind are discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe. Though there never were a circle or triangle in nature, the truths, demonstrated by Euclid, would forever retain their certainty and evidence.

Matters of fact, which are the second objects of human reason, are not ascertained in the same manner; nor is our evidence of their truth, however great, of a like nature with the foregoing. The contrary of every matter of fact is still possible; because it can never imply a contradiction, and is conceived by the mind with the

same facility and distinctness, as if ever so conformable to reality. *That the sun will not rise to-morrow* is no less intelligible a proposition, and implies no more contradiction, than the affirmation, *that it will rise*. We should in vain, therefore, attempt to demonstrate its falsehood. Were it demonstratively false, it would imply a contradiction, and could never be distinctly conceived by the mind.

What Hume is saying can be summarized as follows. There are only two kinds of knowable proposition: relations of ideas, and matters of fact. Relations of ideas have two characteristics: they are either “intuitively or demonstratively certain;” (C1) and they do not assert the existence of any non-abstract entities, such as physical objects or minds or physical or mental events or states (C2).³ Matters of fact, on the other hand, have the following two contrasting characteristics: they do assert the existence of non-abstract entities (C3) and they are neither intuitively nor demonstratively certain (C4). Figure 2.1 expresses these

Relations of ideas	Matters of fact
Are intuitively or demonstratively certain (C1)	Are neither intuitively nor demonstratively certain (C4)
Do not assert the existence of any non-abstract entities (C2)	Do assert the existence of non-abstract entities (C3)

Figure 2.1 Contrasting characteristics of relations of ideas and matters of fact

contrasting characteristics. As we shall see shortly, (C3) requires a qualification: there is a class of propositions which Hume regards as expressing matters of fact even though those propositions do not assert the existence of anything. But before introducing this qualification, an explanation is required of the first three characteristics just mentioned, beginning with (C1) and going counter-clockwise through the above diagram.

Relations of ideas are intuitively or demonstrably certain (C1)

By a proposition which is “intuitively certain,” we may understand Hume to mean one that we can know to be true just by understanding what it says; for example, $1 + 1 = 2$ and a *triangle has three sides*. Notice that the proposition that *I am thinking*, whose certainty was emphasized by Descartes, is not intuitively certain in Hume’s sense, even if we agree with Descartes that this proposition can be known with absolute certainty. For one does not know that this proposition is true just by understanding what it says; rather, one must not only understand the proposition, but also be introspectively aware of one’s own thinking – of a certain psychological fact about oneself. An intuitively certain proposition, as Hume is here

certain,” we shall henceforth say that such a proposition is *self-evident*, because the term “self-evident” nicely brings out the fact that such a proposition makes itself evident or provides the evidence for itself. By a proposition which is “demonstratively certain” we may understand Hume to mean one that follows logically from one or more self-evident propositions.⁵ For convenience, we shall say that such a proposition is “demonstrable.” Relations of ideas, then, are either self-evident or demonstrable.

Relations of ideas do not assert existence (C2)

In saying that relations of ideas do not assert the existence of any non-abstract entities, we are trying to state a key insight of Hume – that which he himself expresses by saying that these propositions are discoverable “without dependence on what is anywhere existent in the universe,” and that “though there never were a circle or triangle in nature, the truths, demonstrated by Euclid, would forever retain their certainty and evidence.” To grasp Hume’s insight, consider his example of the Pythagorean theorem: “the square of the hypotenuse of a right triangle is equal to the squares of its other two sides.” Suppose that no right triangles existed anywhere in the physical world, and that no one had ever even thought about a right triangle. Then the proposition that the square of the hypotenuse of a right triangle is equal to the squares of its sides *would still be true*; for in order for that proposition to be true, there do not actually have to be any right triangles (or even any triangles at all) in the world, nor need anyone ever have thought about a right triangle. The proposition only states a condition that would have to be satisfied by any right triangle that happened to exist or that anyone happened to think about; it just says that if X is a right triangle, then the square on X’s hypotenuse must be equal to the squares on its sides. The basic point is this: propositions like the Pythagorean theorem, or that *three times five is equal to the half of thirty*, remain true regardless of what non-abstract entities exist or do not exist. But then it follows that the proposition does not assert the existence of any non-abstract entities. For if it did, then its truth would depend on whether certain non-abstract entities existed, and so it would not remain true even if those entities did not exist. A proposition that asserts the existence of X cannot remain true regardless of whether or not X exists. The argument:

p remains true regardless of whether X exists or not

∴ *p* does not assert the existence of X

is incontrovertibly valid. For brevity, we shall express Hume’s point by saying that relations of ideas do not “assert existence.” So, in what follows, the phrase “asserts existence” should always be understood simply as an abbreviation for “asserts the existence of some non-abstract entity.”

Matters of fact do assert existence (C3)

In contrast to relations of ideas, propositions expressing matters of fact do – with one important exception, to be noted in a moment – assert existence. Admittedly, Hume does not explicitly state (C3) in the first two paragraphs of Section IV of the *Enquiry* (quoted above), where he presents his “Fork.” Instead, the point about matters of fact that he emphasizes is that “the contrary of every matter of fact is still possible; because it can never imply a contradiction” (E:25; S:15; F:71). However, as we shall see, that point is nothing but a concise statement of Hume’s fundamental argument for (C4)’s claim that no matter of fact is “demonstratively certain.”⁶ (C3), on the other hand, is required in order for (C2) – which, we have seen, Hume does explicitly state, if only briefly and somewhat allusively, in his remarks about mathematics in the first paragraph of Section IV – to be a characteristic that *differentiates* relations of ideas from matters of fact. Furthermore, there is ample textual support, in the rest of Section IV of the *Enquiry* and beyond, for holding that Hume regards (C3) as a defining characteristic of matters of fact: after his initial presentation of the “Fork” in the first two paragraphs of Section IV, he seldom uses the phrase “matter of fact” by itself. Instead, he often uses such phrases as: “*real existence* and matter of fact” (E:26–27; S:16–17; F:72–3); “matter of fact and *existence*” (E:35, E:156, 163–4; S:22, 107, 113; F:79, 187, 193); “matter of fact and *real existence*” (E:35; S:22, F:79); “matter of fact *or real existence*” (E:46; S:30, F:89); “fact *and existence*” (E:158; S:109; F:189; emphasis added). Consider also Hume’s very first example of a matter of fact: “the sun will rise tomorrow.” This proposition asserts the (future) occurrence of a physical event: the sun’s rising tomorrow. And since the proposition cannot be true unless the sun exists, we may also say that the proposition asserts the existence of a physical object: the sun.

There is, however, one kind of proposition that Hume regards as expressing matters of fact even though it does not assert existence. This is any proposition that, taken together with a proposition(s) reporting what is observed at a given time t , or at a set of times t_1, t_2, \dots, t_n , implies the existence of some non-abstract entity which need not be observed at t , or at any of t_1, t_2, \dots, t_n . For example, consider a proposition in which Hume was very interested, namely the so-called “causal principle” or “causal maxim,” which says that *every event has a cause*. Taken by itself, this proposition does not assert the existence of anything. But, taken together with a proposition reporting some observed event, it implies the existence of a cause of that event – whether or not that cause is also observed. We shall call propositions of this sort *bridging propositions*, since they allow one to infer the existence of some thing unobserved from something observed:

Every event has a cause (bridging proposition).

Event e was observed at t (observation).

∴ Event e had a cause, c (item that need not be observed).

Particular causal laws, and laws of nature in general, are bridging propositions. To take a simple example, from the proposition that water boils at 212 F° at sea level, together with the observation that a certain quantity of water is now boiling at sea level, one can infer the existence of a source of heat, whether or not that source is observed; so the law that water boils at 212 F° at sea level is a bridging proposition. As can be seen from our two examples, bridging propositions can have various degrees of generality. The principle that every event has a cause, ranging as it does over all events, past, present, and future, is much more general than the law about water or any other specific law of nature.

Another highly general bridging proposition in which Hume was especially interested (for reasons discussed in Chapter 3) is the so-called “principle of induction” or “principle of the uniformity of nature,” i.e. the principle that “the future will resemble the past” or that “there is a conformity between the future and the past.” That Hume considered this principle to be a matter of fact is clear from his *Abstract of the Treatise*, where he explicitly says: “All probable arguments are built on the supposition that there is this conformity betwixt the future and the past. . . . This conformity is a *matter of fact*”. (T:651; S:131; F:34).

To see that the Principle of Induction is also a bridging proposition, suppose that we conjoin it with the following report of our observations: “Past flashes of lightning have always been followed by claps of thunder, and there is now a flash of lightning.” Then, assuming that the “conformity” or “resemblance” Hume speaks of is understood, as he intends, to assert that events which in our past experience have always been paired will continue to be paired in the future, the principle of induction, conjoined with the proposition reporting our observations, implies that there will be a thunderclap. Thus:

The future will resemble the past (bridging proposition).

Past lightning flashes have been followed by thunderclaps,

and there is now a flash of lightning (observations made at $t_1, t_2, \dots t_n$).

∴ There will be a thunderclap (the as yet unobserved event).

Hume’s “matters of fact,” then, include bridging propositions as well as propositions that actually assert existence, and his relations of ideas exclude them. For convenience, we shall say that bridging propositions “imply existence.” So, in what follows, the phrase “implies existence” will always be an abbreviation for the longer expression, “does not assert existence, but does imply, in conjunction with a proposition(s) reporting what is observed at t , or at $t_1, t_2 \dots t_n$, the existence of some non-abstract entity that need not be observed at t , or at any of $t_1, t_2 \dots t_n$.” Hume’s “matters of fact,” then, either assert or imply existence; his “relations of ideas” neither assert nor imply existence.

Using these abbreviations, together with the terms “self-evident” and “demonstrable” that we substituted for Hume’s “intuitively certain” and “demonstratively certain,” we can update Figure 2.1 and summarize his overall doctrine this way: We come at last to the key

Relations of ideas	Matters of fact
Are self-evident or demonstrable (C1)	Are neither self-evident nor demonstrable (C4)
Do not assert or imply existence (C2)	Assert or imply existence (C3)

Figure 2.2 Matters of fact and relations of ideas: Hume's doctrine summarized

Matters of fact are neither self-evident nor demonstrable

negative thesis of Hume's doctrine – that no proposition stating any matter of fact is either self-evident or demonstrable. This claim has sweeping implications for philosophy. For if Hume is right, then it follows that the only propositions that are either self-evident or demonstrable are those expressing “relations of ideas” – propositions that do not assert or imply existence. By contrast, no proposition that asserts or implies existence is either self-evident or demonstrable. Therefore, what Rationalist metaphysicians tried to do – to demonstrate the existence of various entities such as God, souls, substance, monads, or even the material world – is impossible. In the famous peroration of the *Enquiry*, Hume trumpets forth the sweeping implications of his “Fork:”

When we run over libraries, persuaded of these principles, what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.

(E:165; S:114; F:195)

What Hume's Fork denies, then, is nothing less than the possibility of using reason to demonstrate the existence of anything either natural or divine.

Here it might be objected that we are exaggerating the implications of Hume's doctrine, because not all the major arguments offered by Rationalist metaphysicians are demonstrations in Hume's sense. For example, Descartes' causal proofs of God's existence in his *Third Meditation*, as well as his proof of the physical world's existence in his *Sixth Meditation*, use premisses describing his ideas of God and his ideas of sensible things, respectively, that are not self-evident in Hume's sense, but are supposed to be certain through introspection. However, those proofs also use bridging propositions (notably certain causal principles) which are not certain through introspection and which, if Hume is right, are neither self-

evident nor demonstrable. So Hume’s Fork does imply – assuming (as Descartes and virtually all other seventeenth- and eighteenth-century philosophers did assume, and as more than just a few contemporary philosophers also assume) that knowledge requires certainty, and that the only sources of certainty are introspection, self-evidence, and demonstration – that those arguments do not yield knowledge of their conclusions.

We shall need, therefore, to look carefully at Hume’s rationale for (C4) – his key negative claim that no matter of fact is either self-evident or demonstrable. This will be done partly in the final of the present chapter, and partly in the next chapter. The purpose of the next two parts of this chapter, however, is to relate Hume’s Fork to the well-known formulation of the more modern Empiricist position on knowledge due to Immanuel Kant. So, if you are interested in immediately considering Hume’s rationale for (C4), or in focusing exclusively on the development of Hume’s own theory of knowledge quite apart from Kant’s formulation of the Humean position, then you may wish to go directly to the material in part 5 of the present chapter and in parts 1–4 of Chapter 3. (Except for some cautionary remarks in the third and fourth paragraphs of part 5 of the present chapter and in notes 18 and 19, this material does not presuppose anything in parts 3 and 4 of the present chapter, although part 5 of Chapter 3 does presuppose familiarity with some basic definitional points about “a priori” – “a posteriori” and “analytic” – “synthetic” that are explained at the beginning of the next part.)

3 A modernized version of Hume’s Fork

As we have just seen, Hume’s Fork poses a basic challenge to the ambitions of Rationalist metaphysicians. In our own century, however, this challenge usually has been formulated in a way rather different from Hume’s. Instead of using Hume’s concepts of relations of ideas and matters of fact, recent Empiricists have formulated Hume’s challenge in the a priori–a posteriori, analytic–synthetic terminology introduced by Immanuel Kant. Their position, the modernized version to the extent that it is directly inspired by Hume’s doctrine, can be formulated in this way:

MV All knowable propositions are either analytic a priori or synthetic a posteriori.

In this part, we shall discuss this modernized version of Hume’s Fork. This will enable us to deepen our understanding of the issue between Rationalist and Empiricist traditions. It will also put us in a position to indicate some subtle differences between Hume’s own empiricism and the contemporary empiricism it has inspired – differences that are commonly overlooked but that need to be noticed in order to assess Hume’s position accurately.

Let us begin by going over Kant’s terminology. Kant offers us two pairs of contrasting terms, “a priori” versus “a posteriori,” and “analytic” versus “synthetic,” that are used to

classify statements or propositions along two different dimensions. The a priori–a posteriori contrast classifies statements in terms of how they can be known; it is an epistemological contrast. The analytic–synthetic contrast classifies statements in terms of what makes them true or false; it is a semantic contrast. Let us look more closely at each of these two contrasts.

As we saw in Chapter 1, an a priori proposition is defined as one that is knowable just by thinking (e.g. “no one can be his or her own parent”), and an a posteriori proposition is defined as one that is knowable only by experience (e.g. “some people are over six feet tall”). To avoid possible misunderstanding, we should note here that “knowable just by thinking” does not mean “not knowable by experience;” for it is possible for an a priori proposition to be known by experience. For example, although mathematical statements are a priori, many of them are complex and so are known just by thinking only by those mathematicians who grasp their proofs, while other people may know them on the basis of experience – by hearing of their truth from mathematicians or reading that they are true in mathematics texts. Rather than meaning “not knowable by experience,” then, the phrase “knowable just by thinking” means, roughly speaking, “knowable without experience.” To see why this is still only roughly right, note that for us to know even a simple a priori statement like $1 + 1 = 2$ or *no one can be his/her own parent* some experience is needed; namely, the experience required to learn the meanings of terms – of “1”, “+”, “2,” “parent,” etc. Thus, if “knowable just by thinking” meant “knowable without *any* experience whatsoever,” then no statements would be a priori, because a statement cannot be known unless it is understood; but it cannot be understood unless its constituent terms are understood, and those terms cannot be understood (at least by human beings) unless their meanings have been learned through various sorts of experience. Thus, a more accurate interpretation of the phrase “knowable just by thinking” is this: “knowable without experience, except for the experience required to learn the meanings of terms.”

By contrast, an a posteriori statement cannot be known in this way; for even after one fully understands (the meanings of the terms in) a statement like *some people are over six feet tall*, one may be completely in the dark as to whether the statement is true or false: only experience can determine this. Note, however, that the phrase “knowable only by experience” also requires a bit of clarification; for it does not mean, as one might think, “knowable just by experience.” This is because at least some a posteriori statements require thinking as well as experience in order to be known; for example, our knowledge of scientific laws rests not only on observations but also on complex inferences or extrapolations from those observations.⁷ Thus, instead of meaning “knowable just by experience,” the phrase “knowable only by experience” means “not knowable without experience (other than or in addition to the experience needed to learn the meanings of terms).”

Although “a priori” and “a posteriori” are used to classify statements, they are essentially epistemological terms, pertaining to the *way* in which those statements are *known*. So, we shall follow the common practice of sometimes using these terms apart from reference to

statements. For example, we shall sometimes speak simply of a priori knowledge, meaning knowledge had just by thinking; and of something being knowable a priori or in an a priori manner, meaning that it is knowable just by thinking. Of course, “knowable just by thinking” here does not exclude whatever experience may be needed to acquire the concepts presupposed by the knowledge in question.

Let us now turn to the analytic–synthetic contrast. To avoid misrepresenting the current philosophical landscape, we should first note that the analytic–synthetic distinction is by no means uncontroversial. Some contemporary philosophers, notably the influential American thinker W. V. O. Quine, question the tenability of the distinction; others defend the distinction.⁸ Our purpose here, however, is to see how the distinction is commonly used to define the issue between Rationalists and Empiricists, and to see how that issue, so defined, relates to Hume’s own “Fork.” Therefore, we shall not enter into the controversy concerning the tenability of the analytic–synthetic distinction, but rather assume that the distinction is tenable and merely expound it as it would be usually understood by those who accept it.

An analytic statement is defined as one which is true solely in virtue of the meanings of its constituent terms.⁹ A standard example is “all bachelors are unmarried men.”¹⁰ Notice that although this statement is not couched in the form of a definition – it does not start with “a bachelor is . . .” or “the term ‘bachelor’ means . . .” – it is actually a definition, since “bachelor” means just “unmarried man.” This is why the statement is true solely in virtue of the meanings of its constituent terms. Definitions, then, are one type of analytic statement.

Another type of analytic statement consists of what we may call “conceptual truths.” These are not definitions, but they are still true in virtue of meaning. An example is “something cannot be both round and square.” Although not a definition of either “round” or “square,” this statement is still true in virtue of those terms’ meanings, or of the *concepts* “round” and “square;” for “round” is defined partly in terms of “having no angles,” while “square” is defined in terms of “rectangular,” which in turn is defined in terms of “having four angles.”

The third and most fundamental type of analytic statement consists of statements that are true in virtue of their logical form. Two examples are: “either it is raining or it is not raining;” and “it is not both raining and not raining.” To see why these two statements are true because of their logical forms, we can extract their respective forms, as follows:

Either p or not p .

Not (p and not p).

It is obvious that any statement having either of these forms, no matter what specific sentence one substitutes for “ p ,” must be true. This is why the two statements about rain, as well as any other statements obtained by substituting a given statement for “ p ” in either form, can be said to be “true because of their logical form.” But why, you may ask, are such

statements analytic? The answer is that, like definitions and conceptual truths, they are true solely in virtue of the meanings of their constituent terms. Specifically, our sample statements are true in virtue of the meanings of the terms “either–or,” “not,” and “and,” – terms that give the statements their logical form and that, in the field of logic, are called “logical connectives.”

There is an important relationship between analyticity (i.e. being an analytic statement) and contradiction: the negation (denial) of the truth of an analytic statement is always a contradiction, and conversely the negation of a contradiction is always an analytic statement. Thus, for example, the negation of “all bachelors are unmarried men” is “some bachelors are not unmarried men,” which is a contradiction because “bachelor” means “unmarried man,” so that the negated statement says that some unmarried men are not unmarried men, which is a contradiction (since it means that some men are both unmarried and not unmarried). Conversely, the negation of “some bachelors are not unmarried men” is “all bachelors are unmarried men,” which is analytic. Similar considerations apply to the other examples we have given. Thus, for instance, the negation of “either p or not p ” is “neither p nor not p ”, meaning the same as “not p and (also) not not p ,” meaning simply “not p and p ,” which is of course a contradiction. Conversely, the negation of “not p and p ” is “not (not p and p),” which is analytic.

To turn to the last of Kant's four terms, “synthetic”: a synthetic statement is simply one whose truth or falsity does *not* depend solely on the meanings of its constituent terms. Three such statements are: “all bachelors are taxpayers;” “there are nine planets;” and “every event has a cause.” None of these is true solely by virtue of its constituent terms, and each of them can be denied without embracing a contradiction. Notice how the first statement, “all bachelors are taxpayers,” contrasts with “all bachelors are unmarried men.” Notice also that the third statement could be turned into an analytic truth by substituting “effect” for “event.” As Hume remarked, “every event has a cause” is to “every effect has a cause” as “every man is married” is to “every husband is married” (T:82; F:50).

Using the four Kantian terms that we have now defined, we can classify all possible knowable propositions. As a preliminary point, notice that since one cannot *know* a false proposition – since “S knows that p ” entails that p is true – we can limit this classification to true propositions.¹¹ We could avoid this limitation by construing “knowable” to mean not just “knowably true,” but “knowably true or knowably false,” and by stipulating that “analytic” applies to any proposition that is true *or false* solely by virtue of meaning, that is, to (what are sometimes called) “analytically false” as well as to “analytically true” propositions. But there is no important advantage to be gained by these maneuvers; so we shall confine the classification to true propositions.

Now, no true proposition can be neither analytic nor synthetic: it must be one or the other – either true solely in virtue of meaning, or true not solely in virtue of meaning. Furthermore, no knowable proposition can be neither a priori nor a posteriori: every such proposition must be either knowable just by thinking (i.e. knowable without experience except for the

KNOWLEDGE (I): “HUME’S FORK”

experiences required to learn the meanings of its constituent terms), or else knowable only by experience (i.e. not knowable without experience in addition to the experiences required to learn the meanings of its constituent terms). Thus, every knowable proposition must be either analytic or synthetic and either a priori or a posteriori. Furthermore, it is clear that no proposition can be both analytic and synthetic or both a priori and a posteriori. Thus, going simply by the permutations allowed by Kant’s four terms, there are just four possible classes of knowable propositions: analytic a priori; synthetic a priori; analytic a posteriori; and synthetic a posteriori. We can represent these four classes as shown in Figure 2.3.

	Analytic	Synthetic
A priori	1	2
A posteriori	3	4

Figure 2.3 The four classes of knowable propositions

We can now formulate the issue between the Rationalist and Empiricist schools in terms of these four classes of propositions, for we can use the classification to express both the points of agreement and the point of disagreement between the two schools. Rationalists and Empiricists agree on three points.

- 1 Both affirm that there are instances of classes 1 and 4. It is generally agreed, for example, that at least definitions and the principles of logic are analytic a priori; and that perceptual judgments and the laws of physics are synthetic a posteriori. In addition, many philosophers believe that mathematics consists of analytic a priori statements; and many would say that there are purely conceptual truths which are properly classified as analytic a priori.
- 2 Rationalists and Empiricists agree that there are no instances of class 3, and thus that all analytic propositions are a priori. This is because if a proposition is analytic – if it is true solely by virtue of the meanings of its constituent terms – then surely experience is not required to determine its truth: that can be known just by understanding what the proposition says.
- 3 Rationalists and Empiricists agree that although important, instances of class 1 are

not informative about reality. Consider for example any particular instance of the statement-form “p or not p” – say the statement “either it is raining or it is not raining.” This statement remains true no matter what the weather is like; so it does not convey any information about the weather conditions. The same point applies to any other analytic statement. The statement that all bachelors are unmarried men, for example, would remain true even if there were no bachelors or men or unmarried people, since its truth is guaranteed by the mere synonymy of the words “bachelor” and “unmarried man.” Thus it remains true no matter what anyone’s marital status is, and so does not convey information about anyone’s marital status. Putting the point more grandly, analytic statements do not convey any information about reality, since they remain true no matter what reality is like. This point is rather similar to, though more general than, one of Hume’s key points about his “relations of ideas.” Relations of ideas, you will recall, do not assert the existence of any non-abstract entities, since they remain true no matter what non-abstract entities (e.g. physical or mental objects, states, or events) exist or do not exist. Similarly, the reason why analytic propositions do not describe or convey any information about reality is that they remain true no matter what reality is actually like.

What then is the disagreement between Rationalist and Empiricist philosophers? This: Rationalists affirm, while Empiricists deny, that there are instances of class 2 – synthetic a priori propositions. To properly understand this disagreement, it is important to be clear as to exactly what is meant by denying that there are synthetic a priori propositions. This denial does not mean that there are no propositions which are synthetic but non-empirical (i.e. synthetic but not a posteriori). Rather, it means that any such proposition would be *unknowable*. In what follows, we shall often say that according to empiricism, there are no synthetic a priori propositions. But this will not mean that no proposition is both synthetic and not knowable by experience, but rather that no proposition is both synthetic and knowable just by thinking. Thus, empiricism, as here understood, implies that any proposition that is synthetic but not knowable by experience, while still a genuine proposition, also would not be knowable just by thinking and so would be utterly unknowable (assuming there are no sources of knowledge other than thinking and experience). As we shall see below, some twentieth-century Empiricists – the Logical Positivists – went even further, and claimed that no genuine proposition can be both synthetic and non-empirical. But the weaker doctrine that any such proposition is unknowable, if correct, is strong enough to ruin the hopes of Rationalist metaphysicians.

To gain a better grasp of why Rationalists are committed to affirming, and Empiricists to denying, that there are synthetic a priori propositions, consider the following five propositions:

- (a) There is a reason or explanation for every matter of fact whatsoever.

- (b) Every event has a cause.
- (c) If there is a property, then there must be a substance to which it belongs.
- (d) Existence is a perfection.
- (e) If I can clearly and distinctly conceive of X as existing apart from Y, then X really can exist without Y, at least by God’s power.

This is simply a sample list of propositions that have been upheld by Rationalists. For example, Descartes holds at least (b)–(e), and uses these propositions in constructing proofs of the existence of God and of his famous dualism of mind and body. As for (a), which is called “the principle of sufficient reason,” it has been upheld by Leibniz (who coined the name), Spinoza, and many other thinkers.

Now let us ask: into which of the three available classes of propositions – analytic a priori, synthetic a priori, and synthetic a posteriori – do (a)–(e) fall? Well, it is clear that none of them is true solely in virtue of the meanings of its constituent terms; so they are not analytic a priori. Therefore, they are either synthetic a priori or synthetic a posteriori. But we can eliminate the latter possibility, for historical as well as philosophical reasons. The historical reason is simply that philosophers who rely on these propositions do not take them to be a posteriori (empirical). Descartes, for example, would certainly not say that (b), (c), (d) or (e) rest on experience. Rather, each of these is supposed to be a “clear and distinct” proposition which is either self-evident or deducible from self-evident propositions. As for (a), philosophers who uphold it have usually taken it to be self-evident to any rational mind.

The fact that philosophers have not taken (a)–(e) to be empirical, however, is not the only reason for denying that they are empirical; there are genuinely philosophical reasons. In the case of (c), (d), and (e), these are fairly obvious. Empirical evidence seems simply irrelevant to their truth or falsity; no *observation*, it seems, could possibly count as evidence for or against them: they are simply not the sort of propositions that could be either supported or disconfirmed by observation.

In the case of (a) and (b), matters are more complex. It would no doubt be possible to regard (a) and (b) as expressing an empirical proposition. But then those propositions could not play the role assigned to them in major Rationalist systems like those of Descartes, Spinoza, or Leibniz. For in those systems, (a) and (b) are typically used to establish the existence of things which in the nature of the case cannot be observed, notably God. God is taken, by philosophers who appeal to (a), to be the reason or explanation for the most all-encompassing of matters of fact, namely the existence of the universe itself. Or God is taken, by philosophers who appeal to (b), to be the unseen “first cause” of all other causes. However, if (a) and (b) are regarded as empirical propositions, then they cannot be used in such a manner. The reason for this can be explained as follows. Principles (a) and (b) have the same logical form as the statement “every substance has a solvent”: just as that statement, as understood in modern logic, says “for every substance, there exists some solvent,” so (a) says “for every matter of fact, there exists some explanation,” and (b) says “for every event,

there exists some cause.” Now such statements, unlike simple empirical generalizations of the form “every A is a B,” are not falsifiable: for example, failure to identify a solvent for a substance cannot be used to show that it has no solvent, because we might simply not have looked long enough or far enough for the solvent. However, since “every substance has a solvent” is an empirical statement, the failure to find a solvent for a substance would provide at least some evidence that it has no solvent: it would *disconfirm*, though it would not *falsify* (i.e. definitely refute), the statement that every substance has a solvent. Likewise, if (a) and (b) are construed as empirical propositions, then the fact that we cannot observe God’s existence or creative activity provides some evidence against, or disconfirms, those propositions. But then they cannot reasonably be used as premisses to argue for that very existence or activity. For, surely, if failure to observe *that p* is a disconfirmation of *q*, and we do fail to observe *that p*, then *q* cannot reasonably be used as a premiss supporting *p*. For example, if failure to observe any solvent for a substance of kind *k* disconfirms the proposition that every substance has a solvent, and we do fail to observe any solvent for substances of kind *k*, then the proposition that every substance has a solvent cannot reasonably be used as support for the claim that there is a solvent for substances of kind *k*. Rather than regarding (a) and (b) as empirical claims, then, a Rationalist would hold that failing to observe a reason for some matter of fact, or a cause for a certain event, are not observations that would disconfirm (a) or (b); for (a) and (b) were not based on observation in the first place.

The upshot of all this is that (a)–(e) are neither analytic a priori nor synthetic a posteriori. So, they must be synthetic a priori (assuming that they are knowable at all).

Why do Empiricists deny that there are such propositions? The reason is that such a proposition would be both informative about reality, and knowable just by thinking. But Empiricists are convinced that any proposition that purports to inform us about reality must be based on experience. They find it mysterious, even incredible, that a proposition could be both informative about reality and knowable just by thinking. So, they conclude that the only propositions that can be known just by thinking are analytic a priori propositions, which convey no information about reality, since their truth depends solely on the meanings of their constituent terms. By contrast, Rationalists believe that some genuinely informative propositions, like (a)–(e), can be known just by thinking – that reason is an autonomous source of knowledge about reality.

It was this claim about the power of human reason that led Immanuel Kant, the first philosopher to formulate the issue between rationalism and empiricism in the way just presented, to assign great importance to his famous question: “How are synthetic a priori judgments possible?” As Kant’s way of putting this question shows, he believed that there are synthetic a priori propositions – which makes him, strictly speaking, a Rationalist. However, Kant also drew a distinction *within* the class of propositions that are synthetic but non-empirical (i.e. synthetic but not knowable by experience) – a distinction, roughly speaking, between those that *apply* to experience (though they are not *based* on experience) and those that do not apply to experience. He then argued, against empiricism, that there are

synthetic a priori propositions of the first sort, and, against rationalism, that there are no synthetic a priori propositions of the second sort. So Kant’s own position, which is called “the Critical Philosophy,” is neither purely Rationalist nor purely Empiricist.

In our own century, some Empiricists, known as the “Logical Positivists,” went even further in their condemnation of synthetic a priori propositions than does the modernized version of Hume’s “Fork”. Instead of saying that any proposition which is both synthetic and non-empirical is unknowable, they maintained that, strictly speaking, there are no such propositions. They held that the sentences that *seem* to express such propositions, and which Rationalists hold to express synthetic a priori propositions, are literally meaningless. Such sentences are grammatically well-formed but senseless combinations of words – gibberish akin to “green pains listen chemically.” To support this radical doctrine, the Logical Positivists invoked their notorious “verifiability criterion of meaning.” This doctrine stated that every meaningful sentence must express either an analytic proposition or a proposition which is verifiable or testable by some empirical procedure.

Today this radical version of empiricism, which was at the forefront of philosophy in the 1930s and 1940s, is defunct. One reason for this is that the Logical Positivists were never able to explain clearly what they meant by “empirically verifiable.” Every attempt they made to do so had the unwelcome result either that sentences that were supposedly unverifiable, for example theological sentences turned out to be verifiable after all, or that certain scientific sentences whose meaningfulness the Positivists were anxious to preserve (e.g. sentences in physics about unobservable particles) turned out to be unverifiable and so meaningless. Another reason for the demise of logical positivism was that philosophers soon realized that the verifiability criterion of meaning *itself* is neither analytic nor empirically verifiable; so that the criterion is, in a quite radical sense, self-refuting: it excludes itself from the class of meaningful sentences, i.e. it implies its own meaningfulness. As we shall see in part 4, an analogous difficulty afflicts the modernized version of Hume’s “Fork” that we have been considering.

4 A critique of the modernized “Fork”

If the modernized version of Hume’s “Fork” is sound, then an argument can easily be constructed to show that the propositions typically upheld by Rationalists are unknowable. This “anti-Rationalist argument” starts from the modernized fork (MV hereafter) itself and goes as follows:

MV: All knowable propositions are either analytic a priori or synthetic a posteriori. Propositions typically upheld by Rationalists, such as (a)–(e) (discussed in the previous part), are neither analytic a priori nor synthetic a posteriori.

∴ Propositions typically upheld by Rationalists are unknowable.

This part will offer a critique of this argument. The critique will focus on the leading principle of the argument, MV itself, and will argue that MV is a self-refuting principle, in a sense of “self-refuting” to be explained presently. Then we will examine the implications of this critique for the original version of Hume’s Fork in terms of “relations of ideas” and “matters of fact.”

Suppose that we ask the following question: is MV *itself* analytic a priori or synthetic a posteriori? If it is neither, the anti-Rationalist argument is ruined. For then the Rationalist can give the following counter-argument, which we will call the “anti-Empiricist rebuttal”:

- (A) If MV is knowable, then MV is either analytic a priori or synthetic a posteriori.
(deduced from MV itself)
- (B) MV is neither analytic a priori nor synthetic a posteriori.

- (C) ∴ MV is unknowable.

The upshot is that if MV of Hume’s Fork is true, then it is unknowable!¹²

An analogy may help to clarify the logic of the situation. Consider statement (D): “All knowable statements are less than five words in length.” Clearly, (D) is such that if (D) is true, then – since (D) itself is more than five words long – it cannot be known. We might say that (D) is an “epistemologically self-refuting” statement – one that, if true, is unknowable. Likewise, according to the above argument, MV is such that, if it is true, then – since MV itself is neither analytic a priori nor synthetic a posteriori – it cannot be known.

Since a proponent of MV is logically committed to premiss (A) in virtue of accepting MV itself, the only way she could refute the anti-Empiricist rebuttal would be to refute premiss (B), by showing either that MV is analytic a priori or that it is synthetic a posteriori. I will now argue that neither of these alternatives is tenable.

Let us consider, first, whether MV is analytic. It is certainly not an obviously analytic truth, like “all bachelors are unmarried” or “triangles are three-sided.” It might be suggested, however, that it is an unobviously analytic truth – that it follows logically from statements that are themselves obviously analytic. Whether this is so depends on whether there exists a valid argument for MV, each premiss of which is obviously analytic.

Can one construct such an argument? There is reason to think that Empiricists would subscribe to the following argument, in which “*p*” ranges over all noncontradictory propositions.¹³

- (1) If *p* is not true solely by virtue of meanings, then *p* is true or false by virtue of some feature of reality.
- (2) If *p* is true or false by virtue of some feature of reality, then *p* is not knowable just by thinking.

From these two premisses, it follows that:

- (3) If p is not true solely by virtue of meanings, then p is not knowable just by thinking.

From (3), it follows that:

- (4) If p is knowable just by thinking, then p is true solely by virtue of meanings.

But, by the definition of an a priori proposition as one that is knowable just by thinking, and of an analytic proposition as one that is true solely by virtue of the meanings of its constituent terms, (4) says the same thing as:

- (5) If p is a priori, then p is analytic.

Of course, (5) itself would already rule out synthetic propositions from the class of a priori propositions. So, if the argument as so far developed shows that (5) is analytic, then Empiricists are already vindicated: they need not establish MV's analyticity in order to secure their position. Nevertheless, it is worth noting that MV can be derived from (5), with the help of three further premisses. Those premisses are:

- (6) If p is analytic, then p is a priori.
 (7) If p is knowable, then either p is a priori or p is empirical.
 (8) Either p is analytic or p is synthetic.

MV can now be deduced from (5)–(8), as follows:

- (9) p is a priori if and only if p is analytic (from (5) & (6)).
 (10) If p is knowable, then p is either analytic or synthetic (from (8)).
 (11) If p is knowable, then p is either analytic a priori or synthetic and empirical (from (7), (9), & (10)).¹⁴

But (11) is logically equivalent to MV; so the derivation of MV is now complete.

But does the derivation show that MV is an unobviously analytic truth? It does so only if each of the five premisses (1), (2), (6), (7), and (8) is obviously analytic. Philosophers who accept the analytic–synthetic and a priori–a posteriori distinctions would presumably hold that (1), (6), (7), and (8) are conceptual truths that can be grasped by anyone who understands the terms involved. Since the analytic–synthetic and a priori–a posteriori distinctions themselves are not at issue here, let us grant, at least for the sake of the argument, that those premisses are obviously analytic. This leaves only one premiss to examine: premiss (2).

Is (2) obviously analytic? This question virtually answers itself: far from being obviously analytic, (2) is synthetic. For surely (2) is not a proposition that is true solely by virtue of the meanings of its constituent terms (or whose denial is self-contradictory). Rather, it

expresses a substantive thesis. For the idea behind (2) is this: when a proposition is true or false because of some feature of reality – when its truth or falsity depends on what reality is like – then that proposition conveys some information (or misinformation) about reality: it makes a claim about reality; but when a proposition makes a claim about reality, then it cannot be known by mere thinking. That, of course, is what Empiricists believe: they hold that when a proposition makes a claim about reality, then it can be known only by experience – that reason (rational thinking) is not an independent source of knowledge about reality. Notice that this does not mean just that thinking alone cannot establish the *existence* of certain things, like God. It means also that thinking alone cannot show that things known to exist by experience, such as persons, have certain general features, such as being only material, or only mental, or both material and mental. Now whether this view about the limitations of human reason is correct or incorrect, it certainly does not appear to be an analytic truth. We may conclude, then, that the above argument fails to show that MV is unobviously analytic.

Accordingly, let us next consider whether MV could be synthetic a posteriori. If it is, then, since it says that all knowable statements are analytic a priori or synthetic a posteriori, it would have the logical form of a simple empirical generalization, like “all crows are black.”¹⁵ But as we have seen, simple empirical generalizations (unlike generalizations of the kind “every substance has a solvent,” which say that the items generalized about are related to some existing item, or involve what logic calls “mixed quantification”) must be empirically falsifiable. It must be possible to describe some observation that would show such a statement to be false. Now, how could one show that MV is false? Well, one would have to show that some synthetic proposition(s) are a priori. There are only two ways in which this could be done. One would be to give some abstract argument showing that there are synthetic a priori propositions. But to give such an argument is obviously not to describe any *observation* that would falsify MV. The other way would be to discover the truth of, or actually come to know, some synthetic a priori proposition. Suppose that such a proposition, which we may call “*p*,” did come to be known by a person, S. Could S’s knowing that *p* empirically falsify MV? Clearly, the mere fact of S’s knowing *p* just by thinking could not by itself empirically falsify MV, for it would then be a case of a priori knowledge and so not an observation. However, it might be argued that S’s knowing *p* could empirically falsify MV, in either of two different ways:

- (1) S might come to know *p* just by thinking and also come to know *that she had thus come to know p*, and this second-level knowledge of a synthetic proposition known a priori would itself be an observation that falsified MV; or
- (2) S might know empirically, perhaps via a testimonial chain that S had good reason to trust, that someone else (perhaps an expert on the matter in question) knew *p* just by thinking, and this “second-hand” knowledge that there was a synthetic proposition known a priori would, again, be an observation that falsified MV.

Neither of these, however, constitutes empirical falsification in the relevant sense. For both of them appeal to observations that S can make only if S herself, or someone else, has come to know that p in a non-observational or a priori way. But for a proposition to be falsifiable in the sense that would show it to be an empirical proposition, it must not be the case that some proposition which must be known in order to make the falsifying observation must itself be known a priori. The falsifying observation must not be parasitic, so to speak, on some item of a priori knowledge. Otherwise, it would be possible to use the falsifiability test to "show" that many a priori propositions are really empirical ones. For consider a proposition that can be known a priori to be false, such as $2 + 3 = 6$. Suppose that either (i) S knows a priori that $2 + 3 \neq 6$, and observes that she knows this a priori, or (ii) for some reason S is unable to do any arithmetic, but comes to know that $2 + 3 \neq 6$ on the basis of testimony. Suppose we regard these observations as empirically falsifying $2 + 3 = 6$. Then $2 + 3 = 6$ must be regarded as an empirical falsehood. Furthermore, since the denial of an empirical statement must itself be empirical, $2 + 3 \neq 6$ must be regarded as an empirical statement. Surely, this is absurd: neither (i) nor (ii) shows that $2 + 3 = 6$ is a merely empirical falsehood, or that $2 + 3 \neq 6$ is an empirical statement. Likewise, neither (1) nor (2) above would show that MV is an empirical statement. We may conclude that MV is not an empirical proposition.

Our conclusion concerning MV, then, is that it is an epistemologically self-refuting proposition. (The same conclusion holds for the simpler (5): it too is synthetic but non-empirical, and so unknowable if true.) Therefore, MV cannot be used to show that there are no synthetic a priori propositions, or that propositions like (a)–(e) are unknowable. Note, however, that it does *not* follow from this that there *are* any synthetic a priori propositions, or that propositions like (a)–(e) are knowable. Whether this is so is a question that our critique of MV leaves open. Furthermore, an Empiricist can still argue that certain specific synthetic propositions are not a priori, or that certain types of synthetic proposition are not a priori. For instance, the Empiricist can still argue that no proposition that asserts or implies existence is a priori. As we shall see in the next part and in Chapter 3, this is in effect what Hume does.

Before concluding this part, however, one important question remains. The modernized fork, we have argued, is self-refuting. But what about Hume's own "Fork" – that all knowable propositions are either relations of ideas or matters of fact? Is it self-refuting too? Interestingly enough, the answer to this question is "no." For although it is sometimes uncritically assumed that Hume's relations of ideas are exactly the same as analytic a priori propositions and that his matters of fact are exactly the same as synthetic a posteriori propositions, there are in fact subtle differences between Hume's own Fork and MV. The most important difference is this: the class of propositions that according to Hume's doctrine can be known only by experience is *narrower* than the class of propositions that can be known only by experience according to MV. Concomitantly, the class of propositions that can be known a priori is potentially wider on Hume's doctrine than on MV. The reason for this difference is that Hume's "matters of fact," as we have seen, include only propositions *that assert or imply existence*. Now it is plausible to hold (at least if the ontological argument for God's

existence is assumed to be unsound) that all propositions which assert existence are synthetic – that no non-abstract entities exist “by definition.”¹⁶ It seems safe to say also that propositions which imply existence, like *every event has a cause* and *there is a reason or explanation for every matter of fact whatsoever*, are synthetic. However, the converse does not hold: not all synthetic propositions assert or imply existence. On the contrary, there are some that do not, for instance “there are no three-winged birds.” Hume’s own Fork, unlike MV, leaves open the possibility that some of these synthetic propositions may be knowable a priori. One such proposition may well be Hume’s Fork itself – the proposition that all knowable propositions are either relations of ideas or matters of fact. For this proposition obviously does not assert or imply the existence of anything. Hume’s Fork would then itself have to fall into the class of relations of ideas, since all of his matters of fact do assert or imply existence, and relations of ideas comprise the only other class of knowable propositions. But, while this result may surprise those who assume that Hume’s relations of ideas are exactly the same as analytic propositions, it does seem acceptable. For although all analytic propositions are relations of ideas, the converse need not hold. Hume’s relations of ideas do not have to be analytic: they need only be knowable a priori and not assert or imply existence.

One may well feel that despite the differences between MV and Hume’s own Fork, it should be possible to formulate Hume’s Fork in Kant’s terminology. This is indeed the case; the formulation, which we shall call MV1, would go as follows:

All knowable propositions are either:

- (a) analytic a priori; or
- (b) synthetic a priori propositions that do not assert or imply existence; or
- (c) synthetic a posteriori propositions that assert or imply existence.

Hume’s “relations of ideas” are (a) and (b); his “matters of fact” are (c). Since MV1 itself falls into class (b), MV1, unlike MV, is not self-refuting.

MV1, however, reveals also (as the example concerning three-winged birds may have already suggested) that there is a class of synthetic a posteriori propositions that obviously do not assert or imply existence; namely, negative existentials like *there are no three-winged birds* and *there are no people with IQs of 300*. Indeed, such propositions *deny* the existence of certain non-abstract entities or, as we may say for short, “deny existence.” In order for Hume’s Fork to allow such propositions to be knowable, they would have to be included among the matters of fact. Hume’s own examples of matters of fact are invariably propositions that assert or imply existence; he neither mentions nor gives examples of propositions that deny existence. But such propositions comfortably fit his descriptive phrase “matter of fact and real existence” and its cognates, and so may presumably be included, together with those which assert or imply existence, among the matters of fact. Such a widening of the class of matters of fact is certainly consistent with the spirit of Hume’s doctrine; for it does not enlarge the class of propositions that can be known a priori beyond what he would have

allowed. It only points to a class of a posteriori propositions that he does not explicitly discuss, due perhaps to an overriding interest in the question of how we know propositions that assert or imply existence.

It seems, then, that the most defensible formulation of Hume’s Fork in Kant’s terminology is MV2:

All knowable propositions are either:

- (a) analytic a priori; or
- (b) synthetic a priori propositions that do not assert or imply existence; or
- (c’) synthetic a posteriori propositions that assert, imply, or deny existence.

The “relations of ideas” would be (a) and (b); the “matters of fact” would be (c’). Possibly, there are negative existential propositions concerning non-abstract entities that are synthetic and non-empirical, though it is hard to think of an example. But if there are any such propositions, then (b) should be expanded to read “synthetic a priori propositions that do not assert, imply, or deny existence.”

The Empiricist position expressed by MV2, unlike MV1, cannot be refuted by pointing out that there are synthetic a posteriori propositions that deny existence. And whether MV2 is correct or incorrect, it is in any case not self-refuting; for MV2 itself fails into class (b). We may conclude that unlike the self-refuting MV and the overly restrictive MV1, MV2, which is equivalent to Hume’s own Fork (now understood as including empirical negative existentials in the class of matters of fact), represents a position that is both self-consistent and not obviously incorrect. Its correctness or incorrectness will be explored in the next part and in Chapter 3.

5 Hume’s denials concerning matters of fact

In this section, we shall analyze part of Hume’s case for his key negative doctrine that no matter of fact is either self-evident or demonstrable. In order to maintain a close connection with Hume’s text, our discussion will proceed in terms of his own terminology rather than Kant’s; though, as the discussion of MV1 and MV2 has shown, it could be recast in Kant’s terminology.

Hume’s full defense of his negative doctrine in the *Enquiry* is really the burden of the whole of Section IV: it involves his entire critique of “all reasonings concerning matter of fact.”¹⁷ His central argument for holding that no proposition that asserts existence is demonstrable, in particular, stems directly from his critique of causal reasoning in paragraphs 3–13 of Section IV, part I. And his argument for holding that no proposition that implies existence, or “bridging proposition,” is demonstrable, is given in the course of his critique of inductive reasoning in Section IV, part II. These texts will be analyzed in Chapter 3.

Nevertheless, going just on what Hume says in the first two paragraphs of Section IV, it is possible to discern:

- (1) his reason for saying that no matter of fact is self-evident;
- (2) his general line of argument for saying that no “bridging” matter of fact proposition is demonstrable;
- (3) the key logical point which underlies both that line of argument and his critique of causal reasoning.

Let us look at these three points now. Why does Hume hold that no matter of fact is self-evident? In the second paragraph of Section IV, Hume writes:

The contrary of every matter of fact is still possible; because it can never imply a contradiction. . . . *That the sun will not rise to-morrow* . . . implies no more contradiction, than the affirmation *that it will rise*. We should in vain, therefore, attempt to demonstrate its falsehood. Were it demonstratively false, it would imply a contradiction . . .

The words “every matter of fact” suggest that Hume’s reason for holding that no matter of fact is self-evident is the same as his reason for holding that none is demonstrable, and that this reason is that there is no matter of fact proposition whose denial implies a contradiction. However, we shall resist this interpretation, for the following reason: the proposed reason for denying that any matter of fact is self-evident would be a dangerous one for Hume to use; indeed its use would be self-defeating. For it would require holding that all self-evident propositions owe their status to the self-contradictoriness of their denials and are, therefore, analytic. A proposition whose denial is self-contradictory is always analytic. It would follow that all demonstrable propositions are also analytic, since analyticity is hereditary with respect to entailment – because if P entails Q and P is analytic, then Q is analytic, too. So all relations of ideas, being either self-evident or demonstrable, would be analytic. But this would make Hume’s own Fork – his thesis that all knowable propositions are either relations of ideas or matters of fact – vulnerable to the same self-refutational objection that was raised against MV in the previous section; for which class of knowable propositions would his Fork itself fall into? For reasons like those given for the non-analyticity of MV, Hume’s thesis is not analytic. So, his Fork could not be a relation of ideas, if these are all analytic. Could it then be a matter of fact? No: for it neither asserts, implies, nor denies existence. The upshot is that if Hume’s reason for holding that matters of fact cannot be self-evident is that their denials are never self-contradictory, then his Fork is neither a relation of ideas nor a matter of fact and, therefore, it implies its own unknowability.

It is worth noting also that the view that Hume’s reason for denying that any matters of

fact are self-evident is the noncontradictory status of their denials is at best only weakly supported by the texts. For although there are many passages where Hume gives noncontradictoriness of the denial as the reason why matters of fact are not demonstrable, an examination of those passages provides little evidence that this is also his considered reason for holding that they are not self-evident. For example, in the passage just quoted, Hume’s focus, despite the word “every,” is clearly on what can and cannot be *demonstrated*.¹⁸

What then is Hume’s reason for denying the self-evidential status of matters of fact? The best argument that can be extracted from what he says, I suggest, is this:

- (1) If p states a matter of fact, then p asserts or implies existence.
 - (2) If p asserts or implies existence, then p is not self-evident.
-
- ∴ (3) If p states a matter of fact, then p is not self-evident.

Although this argument is a simple one, it seems very powerful. Its first premiss is true by definition: it simply spells out what Hume means by a “matter of fact.”¹⁹ The second, and key premiss is very plausible. For, recall what a self-evident proposition is: a proposition that we can know to be true just by understanding what it says. On reflection, it seems that, with one possible exception, no proposition that asserts existence can be known in this way. Obviously, propositions like “rocks exist” or “horses exist” are not self-evident. It is true that some philosophers believe (though Hume himself did not believe) that numbers and/or other abstract entities (such as universals) exist, and that those same philosophers would say that the existence of such things is self-evident. But numbers and other abstract entities, even if they can be said really to exist, do not provide counterexamples to premiss (2), because “asserts existence” in (2) means “asserts the existence of some non-abstract entity.”

Perhaps the most challenging possible counterexample to (2) is the proposition *I exist*. It can be plausibly argued that this proposition is self-evident – that anyone who understands it must also know it to be true. The argument turns on the special function of the pronoun “I,” and runs:

- (1) If S understands “I exist,” then S knows what “I” means.
 - (2) If S knows what “I” means, then S knows that “I” denotes or calls attention to the speaker or thinker.
 - (3) If S knows that “I” denotes or calls attention to the speaker or thinker, then S knows that “I exist” is always a true statement.
-
- ∴ (4) If S understands “I exist,” then S knows that “I exist” is always a true statement.

If this is a sound argument, then it seems to show that “I exist” is self-evident. In that case, the argument purporting to show that no matter of fact is self-evident must be amended – by

adding “unless $p = \text{‘I exist’}$ ” to premiss (2) and to the conclusion. And Hume’s position regarding the self-evidential status of propositions that assert existence must be amended to say that, with the exception of the special proposition “I exist,” no proposition that asserts existence is self-evident. This modification, however, would not seriously damage or affect Hume’s overall position.

What about the “bridging” propositions that imply existence? Reflection seems to show that these too are not self-evident. Consider for example the proposition *every event has a cause*. Can that proposition be known to be true just by understanding what it says? The answer seems to be that it cannot. One can understand what the proposition says, without coming to know that there are no *uncaused* events. Of course, the proposition *every effect has a cause* can be known to be true just by understanding what it says, since it is analytic. But that proposition, unlike *every event has a cause*, is not a bridging proposition. For remember that a bridging proposition “implies existence:” when combined with a proposition reporting some observation(s) made at a particular time or times, it enables us to infer the existence of something that need not be observed at that time or those times or at any time. Thus *every event has a cause* is a bridging proposition, because when combined with *event e was observed at time t*, it enables us to infer the existence of a cause of *e*, which we need not observe at *t* or at any other time. But *every effect has a cause* does not work in the same way. For we do not observe at *t* that something is an *effect*, as opposed to being just an *event*, unless we also observe its cause at *t* – in which case the cause is obviously not something that need not be observed at *t*. Furthermore, it seems clear that bridging propositions other than *every event has a cause*, whether they be particular causal laws or the highly general principle that the future will resemble the past, are also not known just by understanding what they say. The argument that we have attributed to Hume for his claim that no matter of fact is self-evident, then, seems to be a sound one.²⁰

Turning now to the question of whether any matter of fact is demonstrable, we have already seen that the reason Hume gives for denying that matters of fact are demonstrable is that their denials never imply a contradiction. This reason works in a rather different way for each of Hume’s two types of matters of fact – those that assert existence and the “bridging” propositions which imply existence. We are not yet in a position to see how it works for propositions that assert existence, since that part of Hume’s argument is inseparable from his analysis of causal reasoning, which will be considered in Chapter 3. But still going only on what Hume says in the second paragraph of Section IV, we can elicit his general line of argument for holding that no “bridging” propositions are demonstrable. It rests on the following principle, hereafter referred to as “principle (P):”

If p is demonstrable, then there is a set of statements, S , such that (1) the members of S are self-evident, and (2) *not p*, together with S , entails a contradiction.

In the second paragraph of Section IV, Hume may be seen as applying principle (P) to a

simple example. He says, in effect: “let p be the proposition *the sun will rise tomorrow*. Then there is no set S of self-evident propositions, such that S together with *not p* entails a contradiction. Therefore, p is not demonstrable.”

Principle (P) itself is based on two points:

- (i) If p is demonstrable, then p follows logically from some self-evident statement(s).
- (ii) If p follows logically from some statement(s), then affirming those statements while denying p entails a contradiction.

(i) follows directly from the definition of a demonstrable statement as one that logically follows from self-evident premisses. (ii) is the point that lies behind Hume’s statement: “The contrary of every matter of fact is still possible, because it can never imply a contradiction.” Hume is here making a basic point of logic: it is always a contradiction to affirm the premisses and to deny the conclusion of a valid argument.

Since (ii) is the key logical point behind principle (P), and since it will also play an important role in Hume’s critique of causal reasoning, let us consider an example to illustrate it. Consider any valid argument, say an argument of the form called “disjunctive syllogism”:

$$\begin{array}{c} P \text{ or } Q \\ \text{Not } Q \\ \hline \therefore P \end{array}$$

What happens if one affirms the premisses and denies the conclusion? Well, to deny the conclusion, P , is to affirm its negation, namely *not P* . Now since we are affirming both premisses, we are also affirming the conjunction of *not P* with the second premiss, that is, we are also affirming *not P and not Q* . But this contradicts the other premiss, P or Q . Obviously, it is a contradiction to say that at least one of two statements is true (“ P or Q ”), and also to say that both of those statements are false (“*not P and not Q* ”). The same thing always results if one affirms the premisses and denies the conclusion of a valid argument – this always leads to a contradiction.

It should now be clear why principle (P) is true: it follows logically from (i) and (ii). For when p is demonstrable, p logically follows from some set, S , of self-evident statements; but when p follows logically from any statements, to affirm those statements while denying p entails a contradiction; therefore, affirming S while denying p must lead to a contradiction.

To illustrate principle (P) itself, we can use a simple arithmetical example of a proposition that satisfies the principle:

Let p = No number is both odd and even.

Then, where n is any number,

$not\ p = n$ is odd and n is even.

But now there is a set, S , of self-evident statements which, together with $not\ p$, entails a contradiction. That set is the pair:

(S_1) If x is odd, then x is not divisible by 2.

(S_2) If x is even, then x is divisible by 2.

This set together with $not\ p$ entails a contradiction, because it follows from $not\ p$ both that

(a) n is odd,

and that

(b) n is even.

Further, it follows from (S_1) and (a) that

(c) n is not divisible by 2,

and it follows from (S_2) and (b) that

(d) n is divisible by 2.

But (c) and (d) contradict each other. The contradiction results because the argument from (S_1) and (S_2) to p is valid, so that a contradiction must result from affirming (S_1) and (S_2) and denying p . This is also a case where p is demonstrable in Hume's sense, since the statements from which p follows, namely (S_1) and (S_2), are both self-evident.

$$\left[\begin{array}{c} (S_1) \\ (S_2) \\ \hline \therefore p \end{array} \right]$$

Now, Hume's general line of argument for the claim that no "bridging" matter-of-fact proposition is demonstrable goes as follows. Let p be any statement that implies existence. Then there is *no* set, S , of self-evident statements, such that S together with $not\ p$ entails a contradiction. Therefore, by principle (P), p is not demonstrable. In Chapter 3, we shall see how Hume uses principle (P) to show that certain specific bridging propositions are not demonstrable.

HUME'S THEORY OF KNOWLEDGE (II)

Causal reasoning and the problem of induction

1 Introduction

Having drawn his basic distinction between relations of ideas and matters of fact in the first two paragraphs of Section IV of the *Enquiry*, Hume devotes the rest of the section to an examination of our knowledge of matters of fact. It is here that we find his full defense of his thesis that no matter of fact, whether it be one that asserts existence or one that implies existence, is demonstrable. However, Hume does not simply advance a straightforward argument for his thesis in respect of each of these two salient points. Rather, his case for each is embedded in a larger context which involves his seminal analysis of causal reasoning and his critique of induction. His overall argument has a clear sequential structure, which successively defends three main theses:

- T1*: Knowledge of matters of fact that is not based on present perception or on memory is always based on causal (i.e. cause-effect) relations.
- T2*: Causal relations are not knowable a priori, but only by inference from past experience.
- T3*: Inference from past experience cannot be rationally justified.

We shall see that Hume's argument for the claim that no proposition that asserts existence is demonstrable stems from *T1* and especially *T2*, and that his argument for the claim that no proposition that implies existence (no "bridging" proposition) is demonstrable comes in the course of his defense of *T3*.

Section IV is divided into "part I" and "part II." Part I presents the case for theses *T1* and *T2*; Part II presents the argument for *T3*. Today, this latter argument is known as "the problem of induction," because inference from past experience is now usually called "induction" or "inductive inference." (Hume seldom uses the term "induction.") The present discussion is organized as follows. Part I is analyzed in the following in section, and part II in the third

section. In the fourth section, we shall look briefly at the psychological account of induction that Hume gives in Section V, part I, of the *Enquiry*. In the final section, we shall discuss one contemporary response to the problem of induction.

2 Hume's critique of causal reasoning

Some of the factual knowledge a person has at any particular time is based on what she is perceiving at that time, and some on what she remembers having previously perceived. For example, your present knowledge that there is a book before you is based on what you now see, and your knowledge that you did or did not eat breakfast today is based on what you remember. Of course, a philosopher influenced by sceptical arguments concerning the senses and concerning memory would say that what you know on the basis of your current perceptions is only that you *seem* to see a book, and similarly that what you know on the basis of memory is that you seem to remember eating (or skipping) breakfast today. In other words, such a philosopher would narrow down the field of what we know by present perception and by personal recollection to certain psychological facts about ourselves, and claim that the rest of our factual knowledge (if any) must rest on some kind of inference from those facts. But everyone, whether or not they would agree in thus narrowing down knowledge based on present perception and personal recollection, would agree that at best only a small portion of what we usually take ourselves to know is based on present perception and personal recollection. For most of our factual knowledge concerns things that are too remote, either in space or in time, or both, for us to have knowledge about them by either present sense perception or personal recollection.

Now Hume's central question in part I of Section IV is: On what is such knowledge based? As he puts it in the topic sentence of the third paragraph:

It may, therefore, be a subject worthy of curiosity, to enquire what is the nature of that evidence, which assures us of any real existence and matter of fact, beyond the present testimony of our senses, or the records of our memory.

(E:26; S:16; F:72)

We may state Hume's question this way:

Q1: When our knowledge of a matter of fact rests neither on present perception nor on memory, on what does it rest?

Hume's answer comes in the topic sentence of the next paragraph: "All reasonings concerning matter of fact seem to be founded on the relation of Cause and Effect. By means of that relation alone we can go beyond the evidence of our memory and senses" (E:26; S:16; F:72). Hume's answer to Q1, then, is the first of the three theses we identified above:

T1: Knowledge of matters of fact that is not based on present perception or on memory is always based on causal relations.

Unless one is careful about how one interprets the phrase “matter of fact,” *T1* may seem too sweeping. If, for example, one interprets “matter of fact” as meaning any synthetic proposition whatsoever, then it seems that some synthetic propositions not known by present perception or memory might be known without any reliance on causal relations. Propositions describing some very general structural features of reality – such as that reality is wholly material, or partly mental and partly material, or wholly mental – might be of this kind. Hume’s Fork itself would be another case in point, since it is a synthetic proposition, but one that can hardly be based on causal considerations. However, let us take our clue from Hume’s expression, “real existence and matter of fact,” and continue to interpret his “matters of fact” as covering only propositions that assert or imply the existence of some non-abstract entity; notably, those that assert or imply the existence of some physical or mental entity or the occurrence of some physical or mental event. Then what Hume is saying is very plausible. For how, except by reasoning from effects to causes or causes to effects, could one know the existence or occurrence of something that one neither perceives nor remembers having perceived? And how can a proposition imply existence – be a bridging proposition that allows us to infer the existence of something unobserved from something observed – unless that proposition asserts some causal relationship between the observed item and the unobserved one?

It is important not to be misled here by the way knowledge is shared and transmitted, especially in our own day and age. We live in an age of rapid, massive dissemination of information. Today, much of our factual knowledge comes from sophisticated electronic media, not to mention more traditional sources like books, teachers, and parents. So Hume’s claim that this knowledge is acquired by causal reasoning may seem, at first, quaint and hopelessly naive. But to dismiss it on these grounds would be an egregious error. For, in the first place, there is a sense in which knowledge acquired from the media, books, and other people does rest on causal relations. For how do we know that these sources are reliable? What assures us that there is any connection at all between what they say and how things really are? The answer, it seems, is that we assume a vast network of causal connections between how things are and how the media, books and other people represent them to be. Were there no such connections, the only way to acquire reliable information would be by one’s own devices. In the second place, Hume is asking especially about the *original acquisition* of knowledge, rather than about the transmission of knowledge already acquired by someone else. But how can one acquire a *new* piece of factual knowledge, in cases where one cannot rely merely on present observation? (Obviously one could not then rely merely on memory, which allows us to store previously acquired knowledge but not to acquire new knowledge.) The only way, it would seem, is by reasoning from effects to causes, causes to effects, effects to other effects of the same cause, etc. In such cases, we simply have nothing to “go on” except causal relations. And since the transmission of knowledge from one person to

another presupposes that someone originally acquired the knowledge on his or her own, Hume seems to have put his finger on an extremely important truth when he suggests that all knowledge of “real existence and matter of fact” not based on perception or memory is ultimately based on causal reasoning.

Granting then the enormous importance of causal relations for acquiring factual knowledge – so granting the plausibility of Hume's thesis T1 – we can turn to the question Hume raises in the next paragraph of his clearly structured discussion: “If we would satisfy ourselves, therefore, concerning the nature of that evidence, which assures us of matters of fact, we must enquire how we arrive at the knowledge of cause and effect” (E:27; S:17; F:73). In other words:

Q2: How do we acquire knowledge of causal relationships?

Unless we can answer this question, we will not understand the basis of the greater part of our factual knowledge.

Hume's answer comes in the topic sentence of the next paragraph:

I shall venture to affirm, as a general proposition, which admits of no exception, that the knowledge of this relation is not, in any instance, attained by reasonings *a priori*; but arises entirely from experience, when we find, that any particular objects are constantly conjoined with each other.

(E:27; S:17; F:73)

Hume's answer to *Q2*, then, is the second of the three theses identified above:

T2: Causal relations are not knowable *a priori*, but only by inference from past experience.

T2 is a vital component of Hume's theory of knowledge; for it virtually establishes his doctrine that propositions which assert existence are not demonstrable. For if knowledge of those propositions rests on causal relations, but causal relations are knowable only by experience, then clearly those propositions themselves are knowable only by experience and, consequently, cannot be demonstrated.

This argument from T2 to the undemonstrability of propositions asserting existence is important, so let us restate it in a complete and careful way (one that “factors-in” the consideration that our knowledge of *some* propositions which assert existence rests on perception or memory rather than on causal relations).

The argument's first premiss, which is actually based on T1, is:

- (1) Knowledge of propositions that assert existence is based on perception, memory, or causal relations.

The crucial T2 enters the argument as its second premiss:

- (2) Causal relations are knowable only by inference from past experience.

This is just the second, affirmative, clause of T2 itself. It now follows from (1) and (2) that:

- (3) Knowledge of propositions that assert existence is based on perception, memory, or inference from past experience.

Furthermore, we may correctly assert that:

- (4) If knowledge of propositions that assert existence is based on perception, memory, or inference from past experience, then propositions that assert existence are not demonstrable.

This is because a demonstrable proposition is, by definition, one that can be deduced from self-evident premisses. But if a proposition can be known only by sense perception or memory or inference from past experience, then obviously it cannot be deduced from self-evident premisses; it has a basis completely different from any proposition that can be so deduced. In different terms, a proposition that must rest on perception or memory or inference from past experience is a posteriori; while a proposition that can be deduced from self-evident premisses is a priori. From (3) and (4), however, there follows the conclusion:

- (5) Propositions that assert existence are not demonstrable.

If Hume can successfully establish T2, then, he will have shown thereby that one of his two classes of matter-of-fact propositions – namely those that “assert existence” – cannot be demonstrated. Accordingly, we shall now look closely at his defense of T2, which occupies the bulk of part I of Section IV and contains some of his most influential reasoning about causal relations.

Hume’s first reason for holding T2 is this. If the cause (or the effect) is something new in our experience, then we cannot tell, just by examining it, what effect it will have (or what its cause was). For example, Adam could not have told, just by seeing that water is something liquid and transparent, that it could drown him; or by seeing that fire is bright and feeling its warmth, that it could destroy him upon a closer approach. This point, Hume notes, is easily admitted for cases where we remember that the cause was once something new in our experience: we admit, for example, that we could not have known, from seeing for the first time two contiguous marble slabs, that they would be hard to pull apart, but easy to slide apart. The point is admitted also for cases where the cause is something relatively rare or uncommon in our experience: who could tell just from looking at gunpowder that it would

explode, or from looking at a U-shaped piece of metal that it would attract iron filings? Finally, the point is easily admitted for cases where the cause involves some intricate or hidden mechanism. We easily admit, for example, that merely looking at the innards of a watch (or a computer!) gives us no knowledge of what it can do. But, Hume says, we do not so easily admit this point for cases concerning which we remember no time when the things involved were new to us, where those things are not at all rare but quite commonplace, and where they involve no intricate mechanism but are quite simple. In such cases, Hume says:

We are apt to imagine that we could discover these effects by the mere operation of our reason, without experience. [For example,] we fancy, that were we brought on a sudden into this world, we could at first have inferred that one Billiard-ball would communicate motion to another upon impulse; and that we needed not to have waited for the event, in order to pronounce with certainty concerning it.

(E:28; S:18; F:74)

But the truth of the matter, Hume declares, is that his point holds just as strongly for cases where the cause is familiar, commonplace, and simple. Take for instance his case of colliding billiard balls. This is as clear and obvious a case of cause-and-effect as one could desire: perfectly familiar, quite commonplace, and involving no hidden mechanisms. Perhaps for this reason (together with the fact that Hume loved to play billiards!), it is his favorite illustration of causation. Now, he points out, even in the case of billiard balls, we can see on reflection that if we had to predict the effect without relying on past experience, then any prediction would be as good as any other. We could predict equally well, among other possibilities, that both balls would come to rest, or that the struck ball would roll off in any one of indefinitely many different directions. As Hume puts it:

[W]ere we required to pronounce concerning the effect . . . without consulting past observation; after what manner, I beseech you, must the mind proceed in this operation? It must invent or imagine some event, which it ascribes to the object as its effect; and it is plain that this invention must be entirely arbitrary.

(E:29; S:18; F:75)

Hume does not confine himself to the point that we cannot predict effects (or “retrodict” causes) without relying on past experience. To support T2, he also attacks a certain conception of causality – one which underlies the view that causal relations can be known a priori, and which we shall call the “Rationalist conception of causality.” Hume never explicitly formulates this conception or announces that he is attacking it, but both the conception itself and Hume’s objections to it can be extracted from what he does say. The Rationalist conception of causality has three related parts:

- (a) the effect is contained in the cause;

- (b) there is a special “tie” or “connection” that binds the effect to the cause; and
- (c) the cause–effect relationship is identical with, or perhaps a special case of, the premiss–conclusion relationship in a valid deduction: causality is identical with, or a special case of, logical deducibility or entailment.

Let us look at each of these points before seeing why Hume objects to them.

Point (a) reflects a way of thinking about causality that can be traced back to Greek philosophy. Causality was conceived of on the model of conception and birth: the effect must be precontained within the cause, almost as the fetus is contained within the womb. The distinguished historian of ideas A. O. Lovejoy calls this idea the “preformationist assumption about causality,” and describes it as follows:

That “there cannot be more in the effect than there is in the cause” is one of the propositions that men have been readiest to accept as axiomatic; a cause, it has been supposed, does not “account” for its effect, unless the effect is a thing which the eye of reason could somehow discern in the cause, upon a sufficiently thorough analysis.

(Lovejoy 1962: 286)

Descartes, for example, thinks of causality in this way; his main proof of God’s existence, presented in his *Third Meditation*, begins with these words:

Now it is manifest by the natural light that there must be at least as much reality in the efficient and total cause as in the effect of that cause. For where, I ask, could the effect get its reality from, if not from the cause? And how could the cause give it to the effect unless it possessed it?

(Descartes 1984: 28)

Descartes’ reasoning is this:

- (1) The cause must contain the reality of its effect.
 - (2) There are degrees of reality.
-
- ∴ (3) There must be as much reality in the cause as in the effect.

Descartes’ argument contains many further steps, which we need not consider here. The point to notice is simply that the opening premiss of his argument is very close to (a). Descartes does not quite say that the cause must literally contain the effect; he says that the cause must contain the *reality* of the effect. But one, very natural, way to understand this rather unclear claim is to interpret it, as does Lovejoy, as saying that the cause itself must

somehow contain the effect. Thus, Descartes' entire argument for God's existence – one of the most important arguments in his philosophy – rests on (a), or at least on something very close to (a).

Locke too, although he is considered to be the founder of classical empiricism, advances an argument for God's existence that relies on (a) or something very like it. Locke begins by arguing that something must have existed from all eternity, since otherwise what exists now would have been “produced by nothing.” He then argues that this something cannot have been just the material universe itself, because

[W]hatsoever is first of all things must necessarily contain in it and actually have, at least, all the perfections that can ever after exist; nor can it ever give to another any perfection that it hath not, either actually in itself or at least in a higher degree: it necessarily follows that the first eternal being cannot be matter.

(Locke 1975: 624)

Locke's point is that since the “perfections” that now exist include thought and intelligence, the “first eternal being” must itself be a thinking, intelligent one. The basic assumption behind this reasoning is, again, that the cause must contain everything (all the “perfections”) that it will ever produce.

Point (b) of the Rationalist conception of causality is not one that is usually articulated. Nor does Hume clarify exactly what the “tie” or “connection” between cause and effect is supposed to be. Perhaps this is because later, in Section VII of the *Enquiry*, he will apply his Empiricist test for meaning to this notion, with dramatic results that he does not wish to anticipate in Section IV. Nonetheless, it would seem that (b) has a place in even the most rudimentary, unreflective thinking about causes and effects. Most people would probably nod their heads in agreement if asked whether there is some “tie” or “connection” between cause and effect, and be shocked if this were questioned. Hume, as we shall see, is quite prepared to shock us here.

Point (c) of the Rationalist conception was probably accepted, at least implicitly, by many philosophers before Hume; one Rationalist philosopher who explicitly asserts it is Spinoza. In his major work, *Ethics Demonstrated in Geometrical Order*, Spinoza says:

But I think I have shown clearly enough (see [Proposition] 16) that from God's supreme power, or infinite nature, infinitely many things in infinitely many modes, i.e., all things, have necessarily flowed, or always follow, by the same necessity and in the same way as from the nature of a triangle it follows, from eternity to eternity, that its three angles are equal to two right angles.

(Spinoza 1985: 426)

Immediately after his initial proof of this proposition, he sets forth the following “Corollary” of the proposition:

Cor. 1: From this it follows that God is the efficient cause of all things which can fall under an infinite intellect.

(*ibid.*: 425)

Spinoza is saying that, from the proposition *everything follows logically from God's nature*, it follows as a first “corollary” that God is the *cause* of everything. What is the assumption behind his reasoning? It must be that causality is identical with, or at least a special case of, logical deducibility. Of course, if this is true, then by knowing the cause, one can know all of its effects just by thinking – by a priori reasoning – which is the heart of what Hume denies.

Hume attacks separately each part of the Rationalist conception of causality. Against (a), he argues that if the effect were contained in the cause, then it would be possible to find or discern the effect by a careful examination of the cause. But, he says:

The mind can never possibly find the effect in the supposed cause, by the most accurate scrutiny and examination. For the effect is totally different from the cause, and consequently can never be discovered in it. Motion in the second Billiard-ball is a quite distinct event from motion in the first; nor is there anything in the one to suggest the smallest hint of the other. A stone or piece of metal raised into the air, and left without any support, immediately falls: but to consider the matter *a priori*, is there anything we discover in this situation which can beget the idea of a downward, rather than an upward, or any other motion, in the stone or metal? In a word, then, every effect is a distinct event from its cause. It could not, therefore, be discovered in the cause, and the first invention or conception of it, *a priori*, must be entirely arbitrary.

(E:29–30; S:18–19; F:74–5)

Drawing on this passage, we can construct the following argument against (a); the notion that the effect is contained in the cause:

- (1) If (a) were true, then we would be able to discern the effect in the cause.
- (2) The effect is an event totally different from the cause.
- (3) If the effect is an event totally different from the cause, then we cannot discern the effect in the cause.
- (4) We cannot discern the effect in the cause (from (2) & (3)).
- (5) (a) is not true (from (1) & (4)).

The key steps in this argument are (2) and (3). In (2), as in the passage just quoted, both the cause and the effect are described as being “*events*.” This reflects an important insight of Hume: namely, that the true members of a cause–effect relationship are events, rather than objects. Often, our ordinary speech masks this fact; for example, we say that “the rock broke

the window.” Here it almost sounds as if the cause is one object (the rock), and the effect another (the broken window). But, of course, what really happened is that the rock hitting the window caused the window to break. Now *the rock hitting the window* and *the window breaking* are not objects or things; they are events or occurrences. Once this point is understood, the claim made in (3) becomes evident. For, given that a cause and its effect are two distinct events, it is false to say that we can discern or discover the effect in the cause; for example, that we can discern or discover the event of the window breaking in the event of the rock hitting the window, in the way that we could discern or discover chocolates in a box of chocolates, or the chocolates’ caramel fillings in the chocolates. This is not because an event can never be part of or contained in another “larger” event: a battle could be part of a war (though the battle would not then be caused by the war, but rather by certain decisions made by the military commanders, movements of troops, and the like). Nor is it even because the effect is always future to the cause: there are cases of simultaneous causation, as when a man makes a footprint in the sand – here the placing of the foot in the sand (cause) occurs at the same moment as the formation of the footprint (effect). Rather, it just seems obvious on the face of it that an event which is the *effect* of another event (as opposed to being a part of it) cannot be discerned in the event that caused it, cannot be discovered just by examining or scrutinizing the cause-event, no matter how minutely. Therefore, by the argument given above, the effect cannot be contained in the cause.

Of course, there are cases of causality that may seem to fit the “containment” model, such as the case of conception and birth. But in such a case, it is true only in a general or rough sense that the mother “caused” the baby. No biologist studying the process of reproduction would describe what happened in such an inexact way. Rather, what happened is that a complex sequence of events involving the mother caused the event of the baby’s birth, and it makes little if any sense to say that any of those events “contained” their effects. It would seem, then, that Hume is on strong ground in thinking that the “containment” view of causality is an erroneously mythological one.

Hume finds (b) no more plausible than (a). Against the “supposed tie or connexion between the cause and effect, which binds them together, and renders it impossible that any other effect could result from the operation of that cause” (E:29; S:18; F:74–75), he argues that it is perfectly conceivable that a cause might have some totally new and unexpected effect:

When I see, for instance, a Billiard-ball moving in a straight line toward another; even suppose motion in the second ball should by accident be suggested to me, as the result of their contact or impulse; may I not conceive, that a hundred different events might as well follow from that cause? May not both these balls remain at absolute rest? May not the first ball return in a straight line, or leap off from the second in any line or direction? All these suppositions are consistent and conceivable. Why then should we give the preference to one, which is no more consistent or conceivable than the rest? All our reasonings *a priori* will never be able to show us any foundation for this preference.

(E:29–30; S:18–19; F:75)

This is basically a challenge to Hume’s reader: since we can conceive of any number of effects following from a given cause, what justification do we have for thinking that there is a special “tie or connexion” between the cause and the effect? In Section VII of the *Enquiry*, entitled “Of the Idea of Necessary Connexion,” Hume will extend this challenge by asking whether the notion of such a connection even makes sense. Notice that the “supposed tie or connexion” between cause and effect that Hume is calling into question is what would enable us to reason *a priori* from the cause to the effect, and that its absence seems to leave such an inference absolutely groundless. As Hume puts it:

And here it is constantly supposed that there is a connexion between the present fact and that which is inferred from it. Were there nothing to bind them together, the inference would be entirely precarious.

(E:26–7; S:16; F:72)

Of course, if (c) were correct – if the causal relation were identical with, or were a special case of, logical entailment – then this alone would mean that such an inference is justified after all. Against (c), however, Hume uses the basic logical point which underlies his principle (P) (see pp. 58–60). Although that point can be extracted from Section IV (e.g. by putting together what Hume says in the second and the tenth paragraphs), it is most explicitly stated in this passage from Hume’s *Abstract of A Treatise of Human Nature*:

[N]o inference from cause to effect amounts to a demonstration. Of which there is this evident proof. The mind can always *conceive* any effect to follow from any cause, and indeed any event to follow upon another; whatever we *conceive* is possible, at least in a metaphysical sense; but wherever a demonstration takes place the contrary is impossible and implies a contradiction. There is no demonstration, therefore, for any conjunction of cause and effect.

(T:650–1; S:130; F:34)

The key point here is that it is never a *contradiction* to deny that the effect will follow the cause, as would have to be the case if (c) were correct.

To see this more clearly, compare the following two arguments:

(I)

Either it is raining or it is snowing.
It is not snowing.

∴ It is raining.

(II)

There is an A-event (a flash of lightning).

∴ There will be a B-event (a clap of thunder).

(I) is, of course, a deductively valid argument; while (II) is an example of causal reasoning. Now to refute (c), Hume drives a wedge deep between these two types of argument. What is the difference between them? Well, suppose we try to accept the premisses and deny the conclusion of each argument.

To deny the conclusion of (I) is to affirm *It is not raining*. This, taken together with the argument's second premiss (which, by hypothesis, we are accepting as true), yields the affirmation *It is not raining and it is not snowing*. But this affirmation contradicts the argument's other premiss, *Either it is raining or it is snowing* (which we are by hypothesis also accepting as true). In other words, by accepting the argument's premisses and denying its conclusion, we are saying, "Either it is raining or it is snowing, but it is not snowing and it is not raining," which is a contradiction – an utterance of the form *p and not p* – since it says both that at least one of the two statements, "it is raining" and "it is snowing," is true, and also that both statements are false. The same thing always results (as already pointed out in part 5 of Chapter 2) from accepting the premisses and denying the conclusion of any valid deductive argument.

But now, suppose we try accepting the premiss and denying the conclusion of argument II. In other words, suppose we say, *There is an A-event (lightning), but there will not be a B-event (thunder)*. In light of our past experience, this may be foolish – but it is *not* a contradiction! To say, "There is lightning, but there will be no thunder," is not to say anything of the form *p and not p*. Therefore, a causal argument like (II) is not deductively valid. This proves that, contrary to (c), the causal relationship is neither identical with, nor a special case of, logical entailment. Causal relations are fundamentally different from logical relations. This distinction is one of Hume's most important contributions to philosophy.

Here you might object, however, that there are cases where a report of the cause does entail a report of the effect. For example, "Oswald killed Kennedy" entails "Kennedy died." Hume's response would be that this example is a cheat; for the premiss "Oswald killed Kennedy" does not report just the cause: it already reports both the cause and the effect. If the premiss had been "Oswald shot Kennedy," then it would have reported only the cause. But that premiss does not entail that Kennedy died; since it would not be a contradiction to assert "Oswald shot Kennedy, but Kennedy did not die." Hume's point is that a premiss that reports or describes only the cause never entails a conclusion that describes the effect. For exactly the same reason, he would say that a premiss that reports or describes only the effect never entails a conclusion that describes the cause.

In order to complete the analysis of Hume's case for holding that causal relations are never knowable a priori, we must now call attention to one final point. Notice that argument II could easily be converted into a valid argument. We need add only another premiss, namely:

If there is an A-event, then there will be a B-event.

Adding this premiss would obviously convert argument II into a valid case of *Modus*

Ponens. Indeed, it seems quite clear that this premiss *should* be added; for our reason for inferring thunder when we see lightning is precisely our belief in such a premiss. Does adding the premiss then rehabilitate the Rationalist idea that causal relations are knowable a priori – that from our knowledge of a cause, we can reason a priori to its effect? No, it does not. For the new premiss itself is not known a priori. Rather, it is based on past experience – on the observed fact that past A-events have been followed by B-events. Instead of rehabilitating the idea that causal relations are knowable a priori, then, the new premiss opens up a new question: what entitles us to infer that just because past A-events have been followed by B-events, future A-events will be followed by B-events? It is to precisely this question that Hume turns in part II of Section IV.

3 The problem of induction

In part II of Section IV, Hume advances his famous argument for the thesis (T3) that inference from past experience cannot be rationally justified. This argument is really a continuation of the case for T2; the argument of part II dovetails with the argument of part I. So, we shall begin our analysis of part II by summarizing the relevant points from part I.

In support of his denial that causal relations are knowable a priori (and in opposition to point (c) of the Rationalist conception of causality), Hume argued that from a premiss of the form

- (1) There is an A-event (e.g. there is a flash of lightning)

we cannot validly deduce a conclusion of the form:

- (2) There will be a B-event (a clap of thunder).

Furthermore, one cannot show that causal relations are knowable a priori by adding the premiss

- (3) If there is an A-event, then there will be a B-event.

For although the argument from (3) and (1) to (2) is now logically valid, (3) is not known a priori. Rather, (3) is inferred from our experience of A-events being followed by B-events. It is to this inference from experience, accordingly, that Hume now turns his attention. Thus, part II begins with these words:

But we have not yet attained any tolerable satisfaction with regard to the question first proposed. Each solution still gives rise to a new question as difficult as the foregoing, and leads us on to farther enquiries. When it is asked, *What is the nature of all our reasonings concerning matter of fact?* [roughly, Q1 posed on page 62]

the proper answer seems to be, that they are founded on the relation of cause and effect [= T1]. When again it is asked, *What is the foundation of all our reasonings and conclusions concerning that relation?* [= Q2] it may be replied in one word, Experience [= T2]. But if we still carry on our sifting humor, and ask, *What is the foundation of all conclusions [e.g. (3), above] from experience?* this implies a new question, which may be of more difficult solution and explication.

(E:32; S:20; F:77)

What exactly is the inference from experience that Hume is proposing to examine? Well, (3) is inferred, roughly speaking, from a premiss of the sort: (4) Past A-events have always been followed by B-events. This formulation is rough, because an accurate report of past experience would say only that *observed* A-events have been followed by B-events, and would have to be restricted even further so as to capture only those cases where the observer remained in a position to observe whether or not a B-event followed the A-event. It would have to say something more like this: "All A-events observed in conditions C, and in cases in which we observed also whether or not a B-event followed, were followed by B-events," where conditions C would vary with the type of events in question. But for simplicity's sake, let us regard the inference from experience that Hume proposes to examine as being the inference from (4) to (3).

Up to this point, then, Hume's analysis of a typical or model case of causal reasoning looks like this:

- (4) Past A-events have always been followed by B-events.
- (3) If there is an A-event, then there will be a B-event (from (4)).
- (1) There is an A-event – a flash of lightning.
- (2) There will be a B-event – a clap of thunder (from (3) & (1)).

Notice that the *numbering* of these statements corresponds to the order in which they would naturally enter into our reasoning: we begin with the observation of an A-event (1), and from this observation we predict a B-event (2); this prediction, however, is supported by our belief that if an A-event occurs then a B-event will follow (3), and that belief in turn is based on our knowledge that past A-events have been followed by B-events (4). Notice also that an alternative wording of (3) (one suggested by Hume's own manner of expression in part II of Section IV) would be: "Future A-events will be followed by B-events." We have used an "if-then" format for (3) in order to preserve the formal validity of the argument from (3) and (1) to (2), which the alternative wording would not do. An amended version of the alternative wording that would preserve formal validity, and which could serve just as well as our formulation of (3), would be: "Present and future A-events will be followed by B-events."

Hume's present question, as we have seen, concerns the inference from (4) to (3): is this

inference from experience legitimate? What, so to speak, are its credentials? His answer comes in these words: “I say then, that, even after we have experience of the operations of cause and effect, our conclusions from that experience are *not* founded on reasoning, or any process of the understanding” (E:32; S:21; F:77). Here Hume is announcing T3 – his famous thesis that inferences from past experience cannot be rationally justified.

Hume’s manner of expressing T3 calls for some special comment. He expresses it in psychological language: instead of saying that conclusions from experience cannot be rationally justified, he says that they are not “founded on reasoning, or any process of the understanding.” This may make it sound as though his point were merely the psychological one that conclusions from experience are not the product of a particular faculty of our minds and of its workings, namely the “understanding” and its “processes.” Now, this is indeed part of Hume’s point. For, as we shall see, Hume goes on to hold, in Section V of the *Enquiry*, that conclusions from experience are the product of a purely psychological principle of association which he calls “custom;” his claim in Section IV that such conclusions are not arrived at by “reasoning, or any process of the understanding,” is meant to prepare the way for his view, propounded in Section V, that they are, instead, the product of “custom.” That view itself is part and parcel of an elaborate psychological theory that is prominent in the *Treatise* but still present, albeit in a muted form, in the *Enquiry*. Furthermore, this psychological theory reflects one of Hume’s chief ambitions, which was to give a comprehensive psychology or “science of MAN” (as he called it in the Introduction of the *Treatise*). He hoped that his psychology would add as much to our knowledge of human nature as Newton, for whom Hume had the greatest admiration and reverence, had added to our knowledge of the physical world; he hoped to be, so to speak, the Newton of psychology. This ambition helps to explain the title of Hume’s first work: *A Treatise of Human Nature*.

Some recent scholars have emphasized this constructive, naturalistic side of Hume’s thought, and de-emphasized its sceptical side. They believe that Hume was more interested in replacing Rationalist ways of thinking about knowledge with a new naturalistic epistemology than he was in defending scepticism.¹ Other, more “traditional,” interpreters of Hume see him as essentially a sceptic.² We shall not try to do justice to this controversy about Hume’s overall intentions; for this would require an extended discussion of the secondary literature.³ Rather, we shall simply maintain that *at least part* of Hume’s purpose in Section IV is to establish that inference from past experience cannot be rationally justified. For his overall argument, as we shall see, is that there is a principle on which all such inferences depend, but that this principle cannot be established or supported by any reasoning or argument. In the *Abstract* of the *Treatise*, Hume sums up the results of his examination of inference from past experience by saying that all such inferences rest on “a supposition . . . which can admit of no proof at all, and which we take for granted without any proof (T:652; S:131; F:34–5).” Thus, both the overall character of his argument and the language he uses to describe it imply that Hume is not totally unconcerned with the question of the rational justification of these inferences – at least part of his message is that such inferences cannot be rationally justified. In the final section of this chapter, I shall argue that, appearances to the contrary

notwithstanding, this negative thesis does not imply that the inferences are *irrational* or *unreasonable* – a point which may reduce the motivation to deny that Hume is even addressing questions about rational justification. But be that as it may, we shall maintain that despite Hume's interest in describing the workings of the human mind, in Section IV part II he is not merely rejecting what he takes to be a mistaken view about how our minds work: he is also putting forward a normative epistemological claim, namely, his thesis, T3, that the inference from (4) to (3) cannot be rationally justified.

Hume's argument for T3 has the following general structure: first, he argues that the inference from (4) to (3) is not valid as it stands; second, he argues that the additional premiss which would make it valid cannot be established. This structure can be seen in the following passage:

These two propositions are far from being the same: *I have found that such an object has always been attended* with such an effect [which is essentially the same as (4)], and *I foresee, that other objects, which are, in appearance, similar, will be attended with similar effects* [essentially the same as (3)]. I shall allow, if you please, that the one proposition may be justly inferred from the other: I know, in fact, that it always is inferred. But if you insist, that the inference is made by a chain of reasoning, I desire you to produce that reasoning. The connexion between these propositions is not intuitive. There is required a medium, which may enable the mind to draw such an inference, if indeed it be drawn by reasoning and argument. What that medium is, I must confess, passes my comprehension; and it is incumbent on those to produce it, who assert, that it really exists, and is the origin of all our conclusions concerning matter of fact.

(E:34; S:22; F:78–9)

Hume's remark that (3) may be "justly inferred" from (4) is ironic. His real point is that there is no valid immediate (one-step) inference from (4) to (3); or, as he puts it, the connection between them is not "intuitive." For, obviously, it would be possible for (4) to be true and (3) to be false; it would not be a contradiction to affirm (4) and to deny (3). So, the only way to get from (4) to (3) would be by way of an additional premiss; or, as he puts it, a "medium" is required to get from one proposition to the other. In saying "what that medium is . . . passes my comprehension," Hume is not saying that he doesn't know what the additional premiss is. On the contrary, he will shortly produce it. His point, rather, is that this additional premiss cannot be established. We shall see presently what the premiss is and why Hume thinks it cannot be established.

First, we should pause to note that Hume often formulates the inference from experience in a different way. Instead of talking merely about something being followed by or "attended with" an effect, he talks about certain sensible qualities possessing certain "secret powers." For example, he points out that just because things possessing the sensible qualities of bread

have had the “secret power” to nourish humans in the past, we infer that such things will have this power in the future. The inference he is criticizing, reformulated in these terms, would go like this:

(4a) Sensible qualities, Q, have had “secret powers,” P, in the past.

∴ (3a) Q will have P in the future.

Hume does not really think that (4a) and (3a) are better formulations than are (4) and (3); on the contrary, he thinks they are worse. For they employ the very notion of a “power” that Hume will attack in Section VII of the *Enquiry*. This is why he adds the footnote to the sentence that introduces the talk of sensible qualities having “secret powers:” “The word, Power, is here used in a loose and popular sense. The more accurate explication of it would give additional evidence to this argument. See Sect. 7” (E:33n; S:21; F:78).

Nevertheless, it is important to appreciate that, as Hume points out, the inference from (4a) to (3a) suffers from exactly the same defect as that from (4) to (3): it is simply invalid. As Hume says, in a passage that closely parallels the passage quoted on page 76:

When a man says, *I have found, in all past instances, such sensible qualities conjoined with such secret powers* [= (4a)]: And when he says, *Similar sensible qualities will always be conjoined with similar secret powers* [= (3a)], he is not guilty of a tautology, nor are these propositions in any respect the same. You say that the one proposition is an inference from the other. But you must confess that the inference is not intuitive; neither is it demonstrative: Of what nature is it, then?
(E:37; S:24; F:81)

The upshot is that bringing in “secret powers,” even if this notion were legitimate, would not salvage the inference from experience. For the fact that certain observed qualities have been joined to certain secret powers in the past would not prove that they will be joined to those powers in the future. Nor, we may add, does the fact that those powers have *operated* in the past prove that they will operate in the future.

Here you might object that Hume’s talk of “secret” powers reflects the backward state of eighteenth-century science, and that the difficulty he is discussing disappears if one thinks in terms of modern science. But this would be to miss the force of Hume’s point; for the inference remains invalid when put this way:

Microstructure S has had power P in the past.

∴ Microstructure S will have power P in the future.

Hume's critique of inference from experience, then, cuts deep; it does not depend on any limitations of the science of his day.

What is the additional premiss or "medium" that would validate the inference from (4) to (3)? Well, there are several possibilities. The simplest would be: "If past A-events have always been followed by B-events, then so will present and future A-events." But instead of this premiss, Hume identifies the most basic and general principle that could play the logical role required to be, as he expresses it, "the supposition, that the future will be conformable to the past" (E:35; S:23; F:80), or that "the future will resemble the past" (E:37; S:24; F:81). Hume is asserting, then, that the premiss that would justify the inference from (4) to (3), and all others like it, is this one:

- (5) The future will resemble the past.

Admittedly (5), like (4), is a rough formulation. In order to deduce (3) from (4), it is not enough to add a vague principle saying that the future will "resemble" or "be conformable to" the past. For that wording allows that the future might not resemble the past closely enough for the inference to go through. Rather, one must interpret (5), quite stringently, as saying that the same patterns of relations between kinds of event – the same laws of nature – will hold in the future as have held in the past. Interpreted in this way, (5) is sometimes called "the principle of the uniformity of nature," and also the "principle of induction." Although the exact interpretation of (5) is a difficult matter that could be discussed at greater length, here we shall simply grant that from (5) – suitably interpreted – together with (4), (3) can be validly derived. At this point, then, Hume's analysis of causal reasoning looks like this:

- (5) The future will resemble the past.
- (4) Past A-events have always been followed by B-events.
- (3) If there is an A-event, then there will be a B-event (alternatively: Present and future A-events will be followed by B-events) (from (4) & (5)).
- (1) There is an A-event (a flash of lightning).
- (2) There will be a B-event (a clap of thunder) (from (1) & (3)).

The obvious next question is whether (5) itself can be safely accepted. Famously, Hume argued that it cannot. His argument is contained in the following passage:

All reasonings may be divided into two kinds, namely, demonstrative reasoning, or that concerning relations of ideas, and moral [probability] reasoning, or that concerning matter of fact and existence. That there are no demonstrative arguments in the case seems evident; since it implies no contradiction that the course of

nature may change, and that an object, seemingly like those which we have experienced, may be attended with different or contrary effects. May I not clearly and distinctly conceive that a body, falling from the clouds, and which, in all other respects, resembles snow, has yet the taste of salt or feeling of fire? Is there any more intelligible proposition than to affirm, that all the trees will flourish in December and January, and decay in May and June? Now whatever is intelligible, and can be distinctly conceived, implies no contradiction, and can never be proved false by any demonstrative argument or abstract reasoning *a priori*.

If we be, therefore, engaged by arguments to put trust in past experience, and make it the standard of our future judgment, these arguments must be probable only, or such as regard matter of fact and real existence, according to the division above mentioned. But that there is no argument of this kind, must appear, if our explication of that species of reasoning be admitted as solid and satisfactory. We have said, that all arguments concerning existence are founded on the relation of cause and effect; that our knowledge of that relation is derived entirely from experience; and that all our experimental conclusions proceed upon the supposition, that the future will be conformable to the past. To endeavour, therefore, the proof of this last supposition by probable arguments, or arguments regarding existence, must be evidently going in a circle, and taking that for granted, which is the very point in question.

(E:35–6; S:22–3; F:79–80)

We can summarize this passage’s overall argument this way:

- Step (i)* There are only two possible ways of establishing a proposition by reasoning: demonstratively, by deductively valid reasoning from self-evident premisses; or inductively, by inductively correct reasoning from rationally acceptable premisses.
- Step (ii)* The proposition (5) “The future will resemble the past” cannot be established demonstratively, because its denial does not imply a contradiction.
- Step (iii)* (5) cannot be established inductively, because all inductive inferences are based on (5), thus rendering an inductive argument for (5) circular.

∴ (5) cannot be established at all.

In what follows, we shall discuss each step of this argument in sequence.

Step (i)

Let us begin with a terminological point. As the wording of (i) indicates, we shall use the term “inductively” to mean “by inductively correct reasoning,” or by reasoning in which the

premisses of an argument render its conclusion probable or likely.⁴ Likewise, the terms “inductive argument” and “inductive inference” will mean “inductively correct argument” and “inductively correct inference,” “inductive” will mean “inductively correct,” and “induction” will refer to inductively correct argumentation in general. This is a slight departure from one common use of those terms. Often, an “inductive argument” is defined as one whose premisses *purport* or are *intended* or *claimed* to provide less than conclusive support for the conclusion. Correspondingly, a “deductive argument” is defined as one whose premisses *purport* or are *intended* or *claimed* to provide conclusive support for the conclusion. Although these definitions are common ones and are practical up to a point, on reflection they are problematic. To see why, consider a couple of examples. A standard example of a “deductive argument” is this:

All humans are mortal.
Socrates is human.

∴ Socrates is mortal.

A typical example of an “inductive argument” is this:

Most adult humans are over four feet tall.
John is an adult human.

∴ John is over four feet tall.

Now suppose that someone gives the first argument, but misguidedly *intends* or *claims* that the premisses provide less than conclusive support for the conclusion, and that someone gives the second argument but misguidedly *intends* or *claims* that the premisses provide conclusive, watertight support for the conclusion. Then it seems that, by the common definitions just mentioned, the first person's argument ought to be classified as an “inductive argument,” and the second person's argument as a “deductive argument.” But people who use those definitions would not accept this result: they would insist that no matter who gives these arguments, and regardless of anyone's intentions, the first argument should be classified as deductive and the second as inductive. Now although it may well be possible to amend the common definitions so as to justify this insistence, this seems by no means an easy task. We can avoid this difficulty by analyzing Hume's position in terms of two types of *logically correct* argument – deductively valid arguments and inductively correct arguments – rather than simply as two types of argument. This is why, throughout the rest of this chapter, we shall use “induction” as a label for inductively correct reasoning, “inductive argument” as short for “inductively correct argument,” and “inductively” as short for “by inductively correct reasoning.”

In light of this terminological point, it might now seem that step (i) ought to be formulated more simply, as follows:

- (i') There are only two kinds of correct reasoning: demonstrative reasoning, and inductively correct reasoning from rationally acceptable premisses.

This formulation would make it look as if Hume's argument starts simply from the division of all logically correct arguments or reasonings into two kinds, as his opening sentence admittedly suggests. Nevertheless, such a formulation would be a poor one. For logically correct arguments do not divide up into demonstrative arguments and inductively correct arguments from rationally acceptable premisses. Rather, they divide up simply into deductively valid arguments and inductively correct arguments. Thus, (i') would be a confused and improper way of saying this:

- (i'') There are only two kinds of correct reasoning: deductively valid reasoning and inductively correct reasoning.

But (i''), while now a correct statement, would not serve Hume's purpose. For he needs to start from a premiss that divides arguments *that can actually be used to establish a thesis* – such as (5), that the future will resemble the past – into two classes. But for that purpose, it will not do to start merely from the standard division of all logically correct arguments into deductively valid and inductively correct arguments. For the mere fact that an argument is deductively valid or inductively correct does not by itself mean that it can be used to establish any thesis. Whether it can also do that depends on whether the *premisses* of the argument are acceptable. Hume shows his awareness of this point by calling one of his two kinds of reasoning “demonstrative”, since such reasoning must not only be deductively valid, but must proceed from self-evident premisses. Our own formulation of his opening premiss as (i) rather than (i'') is intended to allow for the same point, and to do so without the confusion involved in (i'). Of course, the phrase “rationally acceptable premisses” in (i) is vague. However, since Hume is going to argue that (5) cannot be established by *any* inductively correct argument (no matter how credible its premisses), we need not try to eliminate this vagueness.

It might now be objected, however, that (i) is false, because there are ways of arguing for a proposition other than the two allowed by (i).⁵ Suppose that a proposition, *q*, is validly deduced from a proposition, *p*, but that *p* is not self-evident. Then the argument from *p* to *q* is not demonstrative in Hume's sense, since *p* is not self-evident; nor is it merely inductive, since *q* follows deductively from *p*. Hume does not consider this objection, but it seems that he could have answered it as follows. What is the basis of *p*? If *p* is itself deduced from a self-evident proposition(s), then the *complete* argument for *q* – the argument from those self-evident propositions to *p* to *q* – is demonstrative after all. If *p* is not deduced from self-evident premisses, then there are four possible cases to consider.

First, p might be inductively supported by some self-evident proposition(s). Then, in accordance with the principle that an argument can only be as strong as its weakest link, the argument from those propositions to p to q should be classified as *inductive*.

Second, p might be inductively supported by propositions that, while not self-evident in Hume's sense, are known by introspection – propositions like “I seem to see a cat” or “I seem to remember seeing a cat.” Then, in accordance with the same principle, the argument from those propositions to p to q should be classified as *inductive*.

Third, p could be based, either deductively or inductively, on propositions that are neither self-evident in Hume's sense nor knowable by introspection, such as observations of one's physical surroundings (“I see a cat”) or memories (“I saw a cat yesterday”). Then matters become more complicated, but, briefly put, come to this. Insofar as Hume accepts the “way of ideas” doctrine that any statement about the physical world must ultimately rest on introspective reports, and inasmuch as statements about the physical world cannot be simply deduced from such reports but would have to be somehow inferred from them, the argument from introspective reports to p to q would again have to be classified as *inductive*. Alternatively, insofar as Hume holds the sceptical view that statements about the physical world must be inferred from introspective reports but that such an inference cannot succeed, the result would be that neither p nor q can be established at all.

Fourth, the final possible case is where p follows deductively from some statement(s) known by introspection. Terminologically, this would be the most awkward case for Hume, because the argument from those statements to p is then not demonstrative, since introspective reports are not self-evident in Hume's sense; but nor is it inductive, since p follows deductively from the introspective reports. But perhaps Hume could deal with this case by qualifying (i) as follows: the “proposition” in question must not be one that could be known merely by introspection. This qualification would get around the troublesome fourth case, because it seems clear that the only propositions that follow deductively from those that can be known by introspection are other propositions that could be known by introspection; thus, “I seem to see an animal” follows from “I seem to see a cat,” but the former no less than the latter can be known by mere introspection. We shall henceforth assume, then, that the term “proposition” in (i) excludes propositions knowable by mere introspection.

Step (ii)

Hume's reason for denying that (5) is demonstrable is the by-now familiar consideration that its denial does not imply a contradiction. As he says:

May I not clearly and distinctly conceive that a body, falling from the clouds, and which in all other respects, resembles snow, has yet the taste of salt or feeling of fire? Is there any more intelligible proposition than to affirm, that all the trees will flourish in December and January, and decay in May and June? Now, whatever is

intelligible, and can be distinctly conceived, implies no contradiction, and can never be proved false by any demonstrative argument or abstract reasoning *a priori*.

(E:35; S:22; F:79)

In order to analyze this reasoning, we need to recall “principle (P)” which, we saw, underlies Hume’s case for saying that no “bridging” matter-of-fact proposition is demonstrable:

(P) If p is demonstrable, then there is a set of statements, S , such that (1) all the members of S are self-evident, and (2) *not* p , together with S , entails a contradiction.⁶

Early in Section IV Hume pointed out, in effect, that when $p = \textit{the sun will rise tomorrow}$, the condition laid down by this principle is not satisfied; so *the sun will rise tomorrow* cannot be demonstrated. Now, he is asserting that when p is the highly general principle that the future will resemble the past, this condition is still not satisfied; therefore this general principle cannot be demonstrated either. In other words, Hume is saying that (5) cannot be demonstrated, because there is *no* set, S , of statements such that, first, all the members of S are self-evident, and second, *not* (5) together with S entails a contradiction. Putting it differently, he is saying that we cannot, in the following argument, correctly supply any statement(s) to substitute for “ S ”:

not (5): The future will not resemble the past & S (S = some set of self-evident statements).

∴ contradiction

It is very difficult to disagree with Hume about this. For *what* self-evident statement(s), together with “the future will not resemble the past,” entail a contradiction? If there were any such self-evident statements, then it seems that we ought to be able to specify them. For self-evident statements are especially obvious ones, whose truth we can know just by understanding them. Further, the range of relevant statements here is vanishingly small: although there are indefinitely many self-evident statements, such as simple arithmetical truths and obvious analytic truths, they seem logically quite unrelated to proposition (5) or its negation. But it seems clear that we cannot specify any self-evident statements that, together with *not* (5), entail a contradiction. It seems safe to conclude, then, that there are none, and therefore that (5) is not demonstrable.

The reasoning that Hume uses to show that (5) is not demonstrable has an importance that goes even beyond its role as support for T3; for it is this reasoning that enables him, finally, to complete the defense of his “Fork.” To complete that defense, it will be recalled, Hume needed to establish the negative claim we called (C4), that no matter of fact whatsoever,

whether it be one that asserts existence or whether it be a “bridging proposition” that implies existence, is either self-evident or demonstrable (see Chapter 2 part 5). We have already seen Hume’s reasons for holding that no matter of fact of either type is self-evident, and his reasons for saying that no matter of fact which asserts existence is demonstrable; but we still have not fully seen why no bridging proposition is demonstrable.⁷ We have just seen, however, why one important bridging proposition, namely (5), is not demonstrable. Now the reason why no other bridging proposition is demonstrable is exactly the same: Hume’s argument showing that (5) is not demonstrable can be generalized to embrace other bridging propositions.

To see this, consider a more general atemporal analogue of (5): roughly, a principle to the effect that the unobserved resembles the observed. On reflection, it seems clear that this principle does not satisfy principle (P), any more than does (5); for again, there appears to be no set, S, of self-evident propositions which, together with “the unobserved does not resemble the observed,” entails a contradiction. Indeed, Hume’s focus on past-to-future bridging may well be an expository device – a convenient and dramatic way of focusing on *any* principle that would warrant an inference from the observed to the unobserved, whether the unobserved item be past, present, or future. It also seems obvious that no *less* general analogue of (5) – no proposition to the effect that if certain specific kinds of event have been paired in the past, then they will continue to be paired in the future – satisfies principle (P). Finally, in the *Treatise*, Hume argues powerfully that still another bridging principle – the principle that every event must have a cause (or, as Hume puts it, that “every beginning of existence must have a cause of existence”) – does not satisfy principle (P). (T:78–82; F:47–50) It seems very doubtful, therefore, that any bridging principle satisfies principle (P). So, at this point Hume seems finally to have established his doctrine that no matter of fact whatsoever is demonstrable.

Step (iii)

Hume’s reason for denying that (5) can be established inductively constitutes the heart of the classic “problem of induction,” for whose discovery Hume is famous. The passage of the *Enquiry* in which Hume posed the problem was quoted earlier in this chapter, but calls for restatement here:

[A]ll our experimental conclusions proceed upon the supposition that the future will be conformable to the past. To endeavour, therefore, the proof of this last supposition by probable arguments . . . must be evidently going in a circle, and taking that for granted, which is the very point in question.

(E:35–6; S:23; F:80)

Hume makes the same point again a little later:

To say [that the argument for (5)] is experimental, is begging the question. For all inferences from experience suppose, as their foundation, that the future will resemble the past If there be any suspicion, that the course of nature may change, and that the past may be no rule for the future, all experience becomes useless, and can give rise to no inference or conclusion. It is impossible, therefore, that any arguments from experience can prove this resemblance of the past to the future; since all these arguments are founded on the supposition of that resemblance.

(E:37–8; S:24; F:81)

He makes the same point again in the *Abstract* of the *Treatise*:

[W]e could not so much as prove by any *probable* arguments, that the future must be conformable to the past. All probable arguments are built on the supposition, that there is this conformity betwixt the future and the past, and therefore can never prove it. This conformity is a *matter of fact*, and if it must be proved, will admit of no proof but from experience. But our experience in the past can be proof of nothing for the future, but upon a supposition, that there is a resemblance betwixt them. This, therefore, is a point, which can admit of no proof at all, and which we take for granted without any proof.

(T:651–2; S:130–1; F:34–5)

In all these passages, Hume is rehearsing the point, made in step (iii) of his argument, as to why (5) cannot be established: (5) cannot be established inductively because all inductive inferences presuppose (5) as their foundation, thus rendering any inductive argument for (5) circular or question-begging.

To grasp Hume’s charge of circularity, we should first ask: what would an inductive argument for (5) look like?⁸ The easiest way to answer this question is to forget Hume for a moment and simply to ask yourself: “why do I believe that the future will resemble the past?” There is only one possible answer: *because it always has!* The inductive argument for (5), then, must be this:

(6) The future has always resembled the past.

∴ (5) The future will resemble the past.

Now, an inductive argument is one the premisses of which show only that the conclusion is probable, rather than establishing it conclusively. Further, such arguments typically involve reasoning to unobserved cases from a sample of observed cases. The argument from (6) to (5), then, has the hallmarks of an inductive argument: the premiss renders the conclusion only probable (albeit highly probable), and the argument goes from the observed past to the

as-yet unobserved future. Bertrand Russell, in *The Problems of Philosophy* (1912: 65), formulates the argument in a way that nicely captures its inductive character:

(6R) Past futures have resembled the past.

∴ (5R) Future futures will resemble the past.

The argument that Hume rejects as circular or question-begging, then, is the argument from (6) to (5), or from (6R) to (5R). Now, in one sense, Hume's charge of circularity is very puzzling. For what is a circular argument? In the straightforward logical sense, it is an argument that uses its conclusion as a premiss. In the most extreme case, it is an argument with the logical form:

$$\frac{p}{\therefore p}$$

In less extreme cases, it is an argument of the form $q \ \& \ p; \ \text{therefore } p$; or $q \ \& \ p \ \& \ r; \ \text{therefore } p$, etc. But all such arguments are *deductively valid*! For obviously, if p is true as a premiss, p must be true as the conclusion, since the premiss and the conclusion are one and the same statement. A circular argument, then, is automatically a deductively valid argument. But an inductively correct argument cannot be deductively valid: its premisses never entail its conclusion.⁹ It follows that *it is impossible for an inductively correct argument to be circular or question-begging*, at least in the straightforward logical sense. Therefore, the argument from (6) to (5) is not circular in that sense, and Hume would be wrong to say that it is. But nobody thinks that Hume here has committed such a gross, elementary mistake. What, then, does Hume mean when he charges that the argument is “going in a circle” and “begging the question”?

The answer is as follows. The *purpose* of establishing (5) was to *justify all inductive inferences* like the inference from (4), Past A-events have always been followed by B-events, to (3), If there is an A-event, then there will be a B-event. But the inference from (6) to (5) is itself an inductive inference: it is just like the inference from (4) to (3), except for being more general in its scope. So the circularity consists in *using induction in order to justify the use of induction*. Hume's point is not that the inference from (6) to (5) is circular as it stands, or is circular outside the context of his search for a justification of induction. Looked at in that manner, the argument from (6) to (5) is a perfectly respectable inductive argument; indeed (6) is the only possible reason for believing (5). But anyone who uses this argument for the purpose of justifying induction is indeed begging the question, or “taking that for granted which is the very point in question.” For the argument is itself inductive.

Our exposition of Hume's critique of all “reasoning concerning matters of fact” in Section IV part II is now substantially complete. This exposition has been rather complex, so we shall conclude this part by summarizing it.

Hume's critique summarized

It is helpful to think of Hume's discussion as containing both his *analysis* of "all reasonings concerning matter of fact," and his *critique* of that reasoning. His analysis involves the six propositions that we have introduced successively into our exposition. Fully stated, it goes as follows:

- (6) The future has always resembled the past. (Russell's alternative formulation: "Past futures have resembled the past.")
- (5) The future will resemble the past. (Russell's alternative formulation: "Future futures will resemble the past.) [Derived inductively from (6).]
- (4) Past A-events have always been followed by B-events.
- (3) If there is an A-event, then there will be a B-event. (Alternatively: Present and future A-events will be followed by B-events.) [Derived from (4) & (5).]
- (1) There is an A-event (e.g. a flash of lightning).
- (2) There will be a B-event (e.g. a clap of thunder). [Derived from (1) & (3).]

Hume's *critique* of the above reasoning can be summarized virtually in sentence-outline form, as follows:

- I. Causal relations are not knowable a priori, but knowable only by inference from past experience (= T2).

A. (1) is invalid (= Hume's objection to point (c) of the
 ∴ (2) Rationalist conception of causality)

B. (3)
(1) is valid, but (3) is inferred from past experience,
 (2) that is, from (4).

- II. Inferences from past experience (inductive inferences) cannot be rationally justified (= T3).

A. (4) is invalid.
 ∴ (3)

B. The inference remains invalid when formulated in terms of "secret powers:"

(4a) Sensible qualities, Q, have had "secret powers" P in the past.

∴ (3a) Sensible qualities, Q, will have "secret powers" P in the future.

C. The inference from (4) to (3) (or from (4a) to (3a)) becomes valid if (5) is added as a premiss, but (5) cannot be established.

1. There are only two ways of establishing a proposition by argument: demonstratively, by deductively valid reasoning from self-evident premisses; or inductively, by inductively correct reasoning from rationally acceptable premisses.
2. (5) cannot be established demonstratively, because its denial does not imply a contradiction.

Note: More fully, Hume's reasoning here can be put this way:

- (a) If p is demonstrable, then p follows logically from some self-evident statement(s).
- (b) If p follows from any statement(s), then affirming those statements while denying p entails a contradiction.¹⁰
- (c) "Principle (P)": If p is demonstrable, then there is a set, S, of statements such that (1) all the members of S are self-evident, and (2) not p , together with S, entails a contradiction (from (a) & (b)).¹¹
- (d) (5) does not satisfy principle (P) – there is no set, S, of statements such that, first, all the members of S are self-evident and, second, *not* (5) together with S entails a contradiction.
- (e) (5) is not demonstrable (from (c) & (d)).

This reasoning can be generalized to show that no bridging proposition is demonstrable, simply by replacing (d) with:

- (d') No bridging proposition satisfies principle (P) – for any bridging proposition p , there is no set, S, of statements such that, first, all the members of S are self-evident and, second, *not* p , together with S, entails a contradiction;

and replacing (e) with:

- (e') No bridging proposition is demonstrable.

3. (5) cannot be established inductively (i.e. from (6)), because all inductive inferences presuppose (5) "as their foundation," thus rendering an inductive argument for (5) circular.

Note: The inference from (6) to (5) is not circular in the straightforward logical sense: no inductive inference could be. Rather, it is circular because the purpose of establishing

(5) was to justify all inductive inferences, like the one from (4) to (3). But the inference from (6) to (5) is itself an inductive inference, just like (though more general than) the one from (4) to (3). So the circularity consists in using induction in order to justify the use of induction.

Before offering an assessment of Hume's critique, we shall look briefly at the views Hume proposes, in part I of Section V, to replace those which he thinks he has discredited.

4 Hume's psychological explanation of induction

Although Hume is persuaded that inductive inferences cannot be rationally justified, he believes that they can be psychologically explained. It is to this topic that he turns in Section V of the *Enquiry*. That section's title, "Sceptical Solution of these Doubts," can be misleading; especially if one ignores the word "Sceptical." Hume has no intention of arguing that inductive reasoning can be rationally justified after all. Rather, he wants to give a psychological explanation of such reasoning.

You might well ask: why does Hume bother to give this psychological explanation? The general answer is that it fits into his program of giving a complete psychology or "science of man." But perhaps a more specific answer can also be given – one that may go a little way toward explaining why Hume wants to provide such a science in the first place. It is a fact that all sane human beings perform countless inductive inferences every day of their lives. Now suppose someone asked *why* humans make such inferences. One possible answer would be that these inferences are rationally justified. After all, if someone asked why humans accept arguments of the form *Modus Ponens*, one relevant answer would be that such arguments are rationally justified, simply because they are deductively valid. But as regards inductive inferences, this answer is not open to Hume; for he has argued that such inferences cannot be rationally justified. But, then, how does he explain the fact that we all make inductive inferences, and generally arrive at the same results when we do? This is the question Hume tries to answer in Section V, part I. Since he cannot answer it by saying that we make these inferences because they are supported by certain standards of rationality, he says that we make them because our minds are influenced by a psychological principle.

Hume calls this principle "custom" or "habit." By the term "custom," he does not mean some sort of sociological principle. Rather, he means a principle of psychological association. The principle may be put as follows:

If the mind is presented with repeated cases where A is followed by B (e.g. lightning by thunder), it will come to expect B whenever A is presented.

Hume stresses that this is simply a psychological principle that has nothing to do with the

rational or reasoning side of human nature. He stresses also that the sentiment of expecting B or anticipating B, which we experience when A is presented, is something quite definite and strong. In terms of his theory of impressions and ideas, it is a distinctive “impression of reflection.” As we shall see in Chapter 4, this impression plays an important role in Hume’s theory of causality.

Hume has argued that inductive inference cannot be rationally justified; it is the work of a psychological principle that operates independently of logical or rational considerations. Is he saying, therefore, that we should stop making inductive inferences? No, he is not, and it would be a mistake to read him in that way. For, in the first place, Hume does not think that humans have any choice here: we cannot help but form beliefs about the future on the basis of our past experience. As he puts it:

[H]aving found, in many instances, that any two kinds of objects – flame and heat, snow and cold – have always been conjoined together; if flame or snow be presented anew to the senses, the mind is carried by custom to expect heat or cold, and to *believe* that such a quality does exist, and will discover itself upon a nearer approach. This belief is the necessary result of placing the mind in such circumstances. It is an operation of the soul, when we are so situated, as unavoidable as to feel the passion of love, when we receive benefits; or hatred, when we meet with injuries. All these operations are a species of natural instincts, which no reasoning or process of the thought and understanding is able either to produce or to prevent.
(E:46–7; S:30; F:89–90)

In the second place, Hume freely admits – in fact, he insists – that we cannot engage in any goal-oriented action unless we rely on induction. For if we had no beliefs or expectations about future outcomes, we would have nothing to guide our actions. As he puts it:

Custom, then, is the great guide of human life. It is that principle alone which renders our experience useful to us, and makes us expect, for the future, a similar train of events with those which have appeared in the past. Without the influence of custom, we should be entirely ignorant of every matter of fact beyond what is immediately present to the memory and senses. We should never know how to adjust means to ends, or to employ our natural powers in the production of any effect. There would be an end at once of all action, as well as of the chief part of speculation.

(E:44–5; S:29; F:89)

Here Hume declares that all our actions are based on the assumption that the future will resemble the past. He says also that most of natural science (“the chief part of speculation”)

rests on the belief that the course of nature will not change. His view, then, is that while this belief is indispensable both to purposive action and to scientific inquiry, it has no rational foundation. His overall position is therefore this:

- 1 Inductive reasoning cannot be rationally justified.
- 2 Inductive reasoning is instinctual, as well as indispensable to both goal-oriented action and scientific inquiry.

As thinkers (philosophers), we must recognize that our “reasonings concerning matters of fact” are not susceptible of any rational justification. But, as humans planning and acting in the world and inquiring into its workings, we cannot avoid relying on it.

5 Need induction be justified?

Hume, as we have seen, argues that inference from experience, or induction, cannot be rationally justified. In this part, we shall address a question that naturally arises concerning the implications of Hume’s thesis: namely, does it entail that induction is not rational, or irrational? It may seem that the answer to this question must be yes. Furthermore, Hume is often interpreted as holding just such a deeply sceptical view of induction.¹² Although this is a complex matter of interpretation, to which we cannot do justice here, it should be noted that there is certainly reason to doubt that Hume does take such a view. First, as we have seen, he holds that induction is indispensable to purposive action and in science. Second, in some of his writings, such as the *Dialogues concerning Natural Religion*, he holds that some inferences from experience are *better* than others – a view that is obviously incompatible with rejecting all inductive inferences as equally irrational a point well made by Millican (1985: 127–30). Nevertheless, it may seem that, whether or not Hume thinks he has shown induction to be irrational, his critique of induction shows that it is irrational. The question we shall discuss is whether this is really the case. Our thesis will be that it is not – that one can hold that induction cannot be rationally justified *without* holding that it is therefore irrational. In defending this thesis, we shall draw on an incisive discussion of the problem of induction by the Oxford philosopher P. F. Strawson, who is a leading exponent of the so-called “ordinary-language” school of philosophy that flourished at Oxford in the 1950s.

To set the stage for a presentation of Strawson’s position, we need to make two preliminary points. The first is that, in order to show that Hume’s critique of induction entails that induction is irrational, one must rely on a key assumption: namely, that an inductive inference is rational only if it can be supported by some justificatory argument. Simply put, one must assume that *induction is rational only if it can be justified*. (This formulation is not an expression of the truism that induction is rational only if induction is rational; rather, it is a short way of saying that induction is rational only if an argument can be given to justify induction.)

The second preliminary point is that once this key assumption is made, it easily generates scepticism about induction. This is because not just any justification will do: any successful justification of induction must fulfill certain requirements. One crucial requirement is that the justification must not itself rely, at any point, on the use of induction. For of course it would be circular to use induction in order to justify using induction. One cannot justify a species of argument by using that very species of argument. This means that the justification will have to be *deductive* in nature. Indeed, it seems that the justification could consist only in showing that every inductively correct argument can be converted into a deductively valid one.¹³ For otherwise, we would be left with an unsupported inductive component – an “exposed” inductive step, so to speak – that was not backed up by any justificatory argument. Now such a deductive justification of induction is exactly the kind of justification that Hume considers. For, as we saw, he considers the possibility of justifying the inference from

4) Past A-events have always been followed by B-events

to

(3) If there is an A-event, there will be a B-event (or: present and future A-events will continue to be followed by B-events)

by adding the premiss:

(5) The future will resemble the past.

But, as we saw, adding (5) to (4) converts the argument for (3) into a deductively valid one. However, this maneuver can justify the inference from (4) to (3) only if the new premiss, (5), is itself rationally justified. But we cannot justify (5) by giving an inductive argument for it – by inferring it from (6), the statement that the future always has resembled the past. For this would be to fall into circularity; we would be using induction to justify using induction. So, the only course left open is to show that (5) is either self-evident or demonstrable. But clearly (5) is not self-evident: one cannot know it to be true just by understanding what it says. Further, (5) is not demonstrable, since its denial does not imply a contradiction, which is to say, since (5) does not satisfy principle (P).

The upshot is that induction cannot be justified; this much Hume seems to have established. But when this conclusion is combined with the premiss that induction is rational only if it can be justified, the result is obviously scepticism. The way in which Hume's critique of induction is *supposed* to lead to inductive scepticism, then, can be summarized as follows:

(A) Induction is rational only if it can be justified (assumption).

(B) Induction cannot be justified (Hume's discovery).

∴ C. Induction is not rational (inductive scepticism).

In order to see how this position can be criticized, notice that (C) is not the only possible conclusion or moral that can be drawn from (B). A different moral would be this: the assumption that induction is rational only if it can be justified is simply wrong. In other words, perhaps the correct response to the above argument would be to apply to it what we

might call the “pivot principle.” At bottom, a deductively valid argument gives us a choice: either accept the conclusion, or reject at least one premiss. The premisses have no special or privileged status; so, if the conclusion is sufficiently repugnant, we have the option of “pivoting” – of rejecting a premiss. Suppose, for example, that someone were to offer a deductively valid argument purporting to prove that time does not exist – that there is no such thing as time. Our response might be to say: “I know that I had breakfast *before* I had lunch today, so your conclusion that time does not exist is obviously false; therefore at least one of your premisses is false.”¹⁴ Why then, should we not, in the same spirit, reject (A): why should we accept the assumption that induction needs to be justified? Perhaps induction is rational even though it cannot be justified.

For this response to inductive scepticism to be convincing, one needs to build a case against (A) that does not rest just on baldly denying (C). It is here that Strawson’s discussion, in the final chapter of his book *An Introduction to Logical Theory* (1952: 248–63), can help us; for Strawson argues effectively that (A) is false. His basic argument is a simple one: relying on induction is part of what it *means* to be rational, therefore induction is rational even if it cannot be justified.

To support this argument’s premiss, Strawson does two different things. First, he argues that “induction is rational (reasonable)” is an analytic truth. He writes:

It is an analytic proposition that it is reasonable to have a degree of belief in a statement which is proportional to the strength of the evidence in its favour; and it is an analytic proposition . . . that, other things being equal, the evidence for a generalization is strong in proportion as the number of favourable instances, and the variety of circumstances in which they have been found, is great. So to ask whether it is reasonable to place reliance on inductive procedures is like asking whether it is reasonable to proportion the degree of one’s convictions on the strength of the evidence. Doing this is what “being reasonable” means in such a context.

(Strawson 1952: 256–7)

Strawson is here saying that both of the premisses of the following valid argument are analytically true, so that its conclusion too must be analytic:

- (D) It is reasonable to have a degree of belief in p proportional to the strength of the evidence for p .
- (E) The strength of the evidence for p is proportional to the number of favorable instances and the variety of circumstances in which those instances occurred.
-
- ∴ (F) It is reasonable to have a degree of belief in p proportional to the number of favorable instances and the variety of circumstances in which those instances occurred.

But (F) is just a way of saying that induction is reasonable or rational. So, anyone who questions (F) – who asks “is it reasonable to rely on induction?” – is thereby also questioning (D), and asking the absurd question: “is it reasonable to base one’s beliefs on the evidence?”

It might be objected that in giving this argument, Strawson is himself offering a justification of induction. Of course, if any argument that has “induction is reasonable” or some equivalent claim as its conclusion automatically counts as a “justification of induction,” then Strawson is indeed offering his own justification of induction. But this would be a strange way to use the label “justification of induction.” For anyone who harbors sceptical doubts about induction would presumably question (E), or regard (E) as question-begging. For (E) says that there being a large number of favorable instances that have occurred in a variety of circumstances really provides *sound evidence* for *p*: (E) makes a *normative* claim to the effect that inductive evidence is *genuine* or *good* evidence for a proposition. But this is precisely what an inductive sceptic questions. So, the demand for a justification of induction surely *includes* a demand for an argument supporting (E). Strawson’s basic point, however, is that (E) requires no justification, for it is no less an analytic truth than is (D). To bring out (E)’s analyticity, Strawson gives no argument. Rather, he simply reminds us (p. 256) of the way certain phrases are actually used:

Consider the uses of the phrases “good grounds”, “justification”, “reasonable”, &c. Often we say such things as “He has *every justification* for believing that *p*”; “I have *very good reasons* for believing it”; “There are *good grounds* for the view that *q*”; “There is *good evidence* that *r*”. We often talk, in such ways as these, of justification, good grounds or reasons or evidence for certain beliefs. Suppose such a belief were one expressible in the form “Every case of *f* is a case of *g*”. And suppose someone were asked what he meant by saying that he had good grounds or reasons for holding it. I think it would be felt to be a satisfactory answer if he replied: “Well, in all my wide and varied experience I’ve come across innumerable cases of *f* and never a case of *f* which wasn’t a case of *g*”. In saying this, he is clearly claiming to have *inductive* support, *inductive* evidence, of a certain kind, for his belief; and he is also giving a perfectly proper answer to the question, what he meant by saying that he had ample justification, good grounds, good reasons for his belief.

The second thing Strawson does to support the premiss that relying on induction is part of what it means to be rational is to argue that the question, “is induction a justified, or justifiable, procedure?” does not really make sense. He begins by pointing out that it makes perfectly good sense to ask whether *specific* beliefs and *specific* methods of arriving at beliefs are justified, because in asking such questions we are asking how well those beliefs and methods conform to inductive standards. For example, we know what it means to ask

whether the belief that it will not snow in New York State this winter is a reasonable one, or whether the method of looking at tea leaves to predict this year's weather is a rational one. In both cases, we would be asking whether the belief or method is inductively justified: whether past experience supports the belief that it will not snow in New York this winter, or whether it supports the view that looking at tea leaves is a reliable method of forecasting the weather. But when we ask whether an inductive standard or norm itself, such as (E), is reasonable, it is no longer clear what the question means. The question, Strawson suggests, is like asking whether the law is legal. It makes sense to ask whether particular actions are legal. For then we are asking whether they conform to the law. It even makes sense to ask whether particular laws or statutes are legal. For then we are asking whether those laws conform to the highest law of the land (e.g. its Constitution). But it does not make sense to ask whether the highest law of the land is legal (e.g. whether the Constitution is constitutional). For since it is itself the ultimate legal standard, there is no legal standard that it could either conform to or violate. Likewise, when it comes to judging the reasonableness of factual beliefs about the unobserved, and of methods for arriving at such beliefs, induction is itself the ultimate standard or yardstick against which these methods and beliefs are measured. There is no higher standard available by which to assess the reasonableness of inductive standards themselves.

Strawson suspects that some readers will not find his two points convincing. For directly after making them, he says:

It seems . . . that this way of showing the request for a general justification to be absurd is sometimes insufficient to allay the worry that produces it. And to point out that "forming rational opinions about the unobserved on the evidence available" and "assessing the evidence by inductive standards" are phrases which describe the same thing, is more apt to produce irritation than relief. The point is felt to be "merely a verbal" one; and though the point of this protest is itself hard to see, it is clear that something more is required.

(ibid.: 258)

In other words, some people might object that even if we ordinarily *call* inductive evidence "good" or "valid" evidence, and so *call* beliefs based on induction "reasonable" or "rational," and use inductive standards as a yardstick of reasonableness for particular beliefs and methods of inquiry, it does not follow that any of these practices really *are* reasonable or rational. To put it more simply: some people might grant the premiss that relying on induction is part of what it means – part of what *we* mean – by being rational, but refuse to allow that it follows from this mere linguistic fact that relying on induction really *is* rational.

As Strawson's remark that "the point of this protest is hard to see" indicates, he does not really accept this objection. He is prepared to argue from "ordinary language" – from the fact

that we *correctly* call beliefs based on induction “reasonable” – to the claim that those beliefs really *are* reasonable. But instead of just insisting on this basic argument, he makes a further point, which may well be his deepest one. He suggests that the demand for a justification of induction is based on confusing two very different propositions:

- (i) [The universe is such that] induction will continue to be successful.
- (ii) Induction is rational (reasonable).

Proposition (i) is a very general prediction to the effect that induction will continue to yield a rich harvest of true specific predictions about matters of fact. It amounts to saying that the universe will continue to exhibit a sufficient degree of order and regularity for our inductive beliefs about specific matters of fact to be generally true. So, (i) is a *synthetic* proposition: since its truth or falsity turns on how things will be, (i) is not true solely by virtue of its meaning. Further, the only way (i) can be supported is by induction, although this is of course not to justify induction itself; since we would then be using induction. So, (i) is *a posteriori*: it can be known only on the basis of (past) experience. Finally, as Hume saw, (i) could conceivably turn out to be false: the course of nature conceivably could change. Indeed, it is conceivable that nature might become so chaotic that none of our predictions came out true anymore. It is conceivable even that our expectation that we would all perish in such a chaotic universe could turn out false, and that someone could survive to witness this new chaos, in which nothing foreseeable happens. So, (i) is a *contingent* proposition; it is not a necessary truth. In brief, (i) is synthetic, *a posteriori*, and contingent.

Proposition (ii), by contrast, is of a fundamentally different kind. It is analytic, *a priori*, and necessary. In order to show this, Strawson makes (p. 262) perhaps his cleverest point:

The chaotic universe just envisaged . . . is not one in which induction would cease to be rational; it is simply one in which it would be impossible to form rational expectations to the effect that specific things would happen. It might be said that in such a universe it would at least be rational to refrain from forming specific expectations, to expect nothing but irregularities. Just so. But this is itself a higher-order induction: where irregularity is the rule, expect further irregularities. Learning not to count on things is as much learning an inductive lesson as learning what things to count on.

In other words: even in a chaotic universe, induction would not cease to be rational. For now it would be rational not to make any specific predictions; in other words, to predict just more chaos or irregularity. But this prediction would itself be based on induction!

The most important implication of the contrast between (i) and (ii) is this: (ii) does not

depend on (i). Even if (i) became false, (ii) would still be true. This means that showing how (i) conceivably could become false (as Hume does) has no tendency at all to show that induction is not rational.

Let us apply this point directly to the question of whether Hume's critique of induction implies that induction is not rational. Those who believe that it does presumably hold that whether an inductive inference like that from (4) to (3) is rational depends on whether (5) – the principle that the future will resemble the past – can be established. Simply put, they hold that:

(G) (ii) depends on (5).

Now, notice that (i) – the statement that [the universe is such that] induction will continue to be successful – is very closely related to (5) – the statement that the future will resemble the past. The only difference between them is that (i) does, while (5) does not, explicitly say that induction has and will be used; since (5) asserts only the uniformity of nature, while (i) affirms also the occurrence of induction within this uniform nature. Accordingly, consider a slightly modified version of (i):

(ia) [The universe is such that] induction, if correctly used, will continue to be successful.

Then it seems safe to say that (5) is true just in case (ia) is true; in other words, that:

(H) (5) is true if and only if (ia) is true.

Those who think that Hume's position implies the irrationality of induction, then, are committed to the logical consequence of (G) and (H), namely:

(I) (ii) depends on (ia).

But surely (I) is false: just as Strawson's chaotic universe scenario shows that (ii) does not depend on (i), it likewise shows that (ii) does not depend on (ia). So, since (H) is true, (G) must be false. In other words, we may respond to inductive scepticism by reversing or applying the "pivot principle" to the argument from (G) and (H) to (I), as follows:

(Not-I) (ii) does not depend on (ia).

(H) (5) is true if and only if (ia) is true.

∴ (Not-G) (ii) does not depend on (5).

We may conclude, therefore, that in order for inductive inferences like that from (4) to (3) to

be rational, it is not necessary to justify them by means of (5). Therefore, even if we accept Hume's demonstration that it is impossible to justify induction, we need not accept the sceptical conclusion that induction is therefore irrational.¹⁵ Instead, we can reasonably maintain that induction is rational, even though it cannot be justified.¹⁶

HUME'S THEORY OF CAUSALITY

1 Introduction

As we have seen, Hume was deeply interested in the relation of cause and effect. For he held that all factual beliefs, except those based on present perception and on memory, ultimately rest on causal relationships. In Section IV of the *Enquiry*, accordingly, he focused on the question of how we acquire and use causal knowledge, or at least (as those who interpret Hume as an inductive sceptic would say) causal beliefs. In Section VII, entitled "Of the Idea of Necessary Connexion," he returns to the topic of causality, but approaches it from a different point of view. His interest now is no longer primarily epistemological; he is no longer primarily concerned with our *knowledge* of causal relationships. Rather, his interest in Section VII is an analytical or definitional one: he wants to discover what causality is, quite apart from the question of how we know whether or what causal relationships exist in the world.

More specifically, he has two related purposes. First, he wants to explain the meaning of the terms "power," "force," "energy," "necessary connection," and the like. (In his discussion of the same topic in the *Treatise*, he mentions also the terms "efficacy," "agency," "necessity," "connection," and "productive quality.") He wants to understand such terms, because they are obviously implicated in the notion of causality: when a cause operates, it is commonly thought to do so because it then exercises power, force, or energy; it is also thought of as producing or necessitating its effect, which is in that sense necessarily connected with the cause. Second, Hume seeks to give what contemporary philosophers call an "analysis" of causation: a definition that specifies, in a noncircular manner, only and all of those statements that must be true for a causal relationship to obtain, or that give the necessary and sufficient conditions for the statement *X caused Y*.

Throughout Section VII, Hume is guided by the principle of empiricism that he advanced in Section II. Thus, immediately after introducing the topic of necessary connection, he reasserts that principle:

It seems a proposition, which will not admit of much dispute, that all our ideas are

nothing but copies of our impressions, or, in other words, that it is impossible for us to *think* of anything, which we have not antecedently *felt*, either by our external or internal senses.

(E:62; S:41; F:102)

This serves as a strong reminder that everything Hume will say about necessary connection and its cognates will be governed by his principle of empiricism. This principle, accordingly, dictates the method Hume will follow in order “to fix, if possible, the precise meaning of these terms” (E:62; S:40; F:102): he will apply his empirical test for meaning to them. In other words, his chief question about “necessary connexion” and its cognates will be: *from what impression(s) is the idea of necessary connection derived?* As he says:

To be fully acquainted, therefore, with the idea of power or necessary connexion, let us examine its impression; and in order to find the impression with greater certainty, let us search for it in all the sources, from which it may possibly be derived.

(E:63; S:41; F:103)

Having announced his method, Hume launches into a search for the impression(s) of necessary connection. This impression hunt occupies much the greater part of Section VII – all of part I, and the first half of part II. Not until the second half of part II does Hume give, in the space of a single paragraph, his famous analysis of causality. We shall examine his search for the impression of necessary connection in part 2, and his analysis of causality in the third and fourth parts of this chapter.

2 Necessary connection

When looking for an observable item – an “impression” – answering to the idea of necessary connection, the natural way to proceed is by scrutinizing any pair of events related as cause and effect. Now some of the clearest cases of cause and effect are those involving ordinary material objects, such as Hume’s favorite case of colliding billiard balls. It is with such cases, accordingly, that Hume begins his impression hunt. He asks us to consider two colliding billiard balls, and to reflect on the question: what do we actually observe in such a case? Well, we observe the cue ball rolling toward the object ball; we see the cue ball come into contact with the object ball, and we see the object ball roll away from the place where it was struck. No doubt, we also hear a noise when the balls come into contact. But do we see the cue ball exercising a power, or exerting a force, on the object ball? Do we see a connection or tie between the events involved – between the collision of the balls and the ensuing motion of the object ball? Do we even know what such a connection or tie would look like? To all these

questions, Hume's answer is a resounding NO. Furthermore, since the case of the billiard balls is as clear, obvious, and typical a case of causation as one might like, Hume generalizes from it: he concludes that no individual case of causation involving only objects that we perceive by our senses yields any impression of necessary connection. As he puts it:

When we look about us toward external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. The impulse of one billiard-ball is attended with motion in the second. This is the whole that appears to the *outward* senses.

(E:63; S:41; F:103)

In reality, there is no part of matter, that does ever, by its sensible qualities, discover any power or energy, or give us ground to imagine, that it could produce any thing, or be followed by any other object, which we could denominate its effect. Solidity, extension, motion; these qualities are all complete in themselves, and never point out any other event which may result from them. The scenes of the universe are constantly shifting, and one object follows another in an uninterrupted succession; but the power or force, which actuates the whole machine, is entirely concealed from us, and never discovers itself in any of the sensible qualities of body. We know that, in fact, heat is a constant attendant of flame; but what is the connexion between them, we have no room so much as to conjecture or imagine. It is impossible, therefore, that the idea of power can be derived from the contemplation of bodies, in single instances of their operation; because no bodies ever discover any power, which can be the original [origin] of this idea.

(E:63-4; S:42; F:104)

Having argued that the idea of necessary connection cannot be derived from observing any individual event-pair in the physical world around us, Hume considers another possibility. Perhaps the idea of necessary connection is derived from introspection; specifically, from experiencing the operation of our own wills. In other words, perhaps one can detect, by introspection, a necessary connection between one's own will and its effects. Hume examines, at some length, two subpossibilities: (1) that we can detect some necessary connection between our will and its effects when we perform voluntary bodily actions (such as raising an arm); and (2) that we can detect some necessary connection between the individual will (hereafter, the will) and its effects in deliberate thinking and/or imagining (e.g. when we conjure up a mental image). Hume argues against both of these possibilities, using three arguments against each of them. (He labels these arguments "*first*," "*Secondly*," and "*Thirdly*".)

We can summarize his arguments against (1) as follows. First, if we were aware of a

necessary connection between the will and its effects in voluntary action, then we would understand how the mind interacts with the body. Since we do not grasp the “secret union of soul and body,” it follows that we are not aware of any such connection. Second, if we could detect a power or force operating when we exercise the will in order to control our body, then we would understand exactly why we can voluntarily control some parts of our body (e.g. the legs and fingers) but not others (e.g. the liver and spleen). For then we would be aware of this force's presence in the former cases, and of its absence in the latter cases. But, says Hume, “we cannot assign any reason besides experience, for so remarkable a difference between one and the other” (E:65; S:43; F:105). All we know is that, in some cases, willing a certain bodily movement is followed by that movement; while in other cases, the movement does not occur no matter how intensely we may desire it. Third, science teaches us that the immediate effect of the will in voluntary movement is not the intended movement itself. Rather, it is an event of which we are wholly unaware and probably ignorant – presumably a brain event. (Although Hume himself does not here mention brain events, from our twentieth-century perspective it may be helpful to illustrate his point by reference to them.) This brain event in turn causes a series of neurological and physiological events, of which we are equally unaware, until at last the desired movement occurs. Now since we are unaware of the brain event, we are obviously also unaware of any necessary connection between it and the will. Again, all we are aware of is that willing a certain movement is followed by that movement. We are not aware of any necessary connection between volition and the brain event that initiates the sequence of neurological events, muscular contractions, etc., which ultimately leads to the desired movement.

We can summarize Hume's arguments against (2) as follows. First, when we deliberately conjure up an idea or image, we are of course conscious of that idea or image itself. But we have no understanding at all of the power by which we bring it to consciousness. Indeed, Hume suggests that such purely mental causation is even more mysterious than any so far considered, comparing it to God's creation of things *ex nihilo*, out of nothing:

This is a real creation; a production of something out of nothing: which implies a power so great, that it may seem, at first sight, beyond the reach of any being, less than infinite.

(E:68; S:45; F:108)

Volition is surely an act of the mind, with which we are sufficiently acquainted. Reflect upon it. Consider it on all sides. Do you find anything in it like this creative power, by which it raises from nothing a new idea, and with a kind of *Fiat*, imitates the omnipotence of its Maker, if I may be allowed so to speak, who called forth into existence all the various scenes of nature? So far from being conscious of this energy in the will, it requires . . . experience . . . to convince us that such extraordinary effects do ever result from a simple act of volition.

(E:69; S:45; F:108–9)

The upshot is, as before, that while we are conscious of the cause (e.g. willing or deciding to picture a triangle) and of the effect (the image of the triangle), we are not aware of any power or force by which the cause produces the effect. As Hume puts it: "We only feel the event, namely, the existence of an idea, consequent to a command of the will: but the manner, in which this operation is performed, the power, by which it is produced, is entirely beyond our comprehension" (E:68; S:45; F:108).

Hume's second and third arguments against (2) are variants of his second argument against (1). If we were aware of a necessary connection between the will and our thoughts, then we would be able to understand (i) why we have far less control over some of our thoughts (e.g. feelings like envy, anger, elation) than others (e.g. plans for tomorrow) and a rather limited control over them on the whole, and (ii) why we have more control over our thoughts at some times (e.g. when sober) than at other times (e.g. when intoxicated). But (second argument) we cannot "assign the ultimate reason of these boundaries, or show why the power is deficient in one case and not in another" (E:68; S:45; F:108); nor (third argument) can we "give any reason for these variations, except experience" (E:68; S:45; F:108). These two arguments cannot be refuted simply by giving psychological or biological explanations of our weak and/or variable control over our thoughts and emotions. For these explanations would surely appeal to "experience" – to empirical generalizations such as "when a man consumes alcohol, his control over his thoughts diminishes." But Hume's point is precisely that such explanations based on experience are the *only* ones possible; whereas if we could detect a necessary connection between our will and our thoughts, then we could explain (i) and (ii) simply in terms of that connection itself, *without* referring to past experience or generalizations inferred from it.

There is a possible objection to Hume's arguments against (1) and (2) that we should consider. This is that in all these arguments Hume makes an unwarranted assumption. He assumes that in order to be aware *of* some connection between the will and its effects, we must also grasp or understand *how* this connection works. This assumption is very explicit, for example, in this passage from his third argument against (1): "[I]f the original power were felt, it must be known: Were it known, its effect must also be known; since all power is relative to its effect. And *vice-versa*, if the effect be not known, the power cannot be known nor felt." (E:66; S:44; F:106). Here Hume is saying that in order even to *feel* a power, one must "know" it, apparently in the sense of being able to foretell its effect without relying on past experience. In other words, he seems to be assuming that only a power or connection that would license a deductive inference from cause to effect or effect to cause – an inference of the sort he rejected in Section IV – could qualify as an impression of power or necessary connection. The same assumption was operative earlier, when he argued that we observe no necessary connection between cause and effect in the world around us:

When we look about us toward external objects . . . we are never able, in a single instance, to discover any power or necessary connexion, any quality which *binds*

the effect to the cause, and renders the one an infallible consequence of the other.
(E:63; S:41; F:103; emphasis added)

Here again, Hume seems to be assuming that any power or necessary connection of which we might be aware must be one that makes the effect deducible from (“an infallible consequence” of) the cause. But this assumption is questionable. Suppose, for example, that you force yourself to hold up a heavy object, or to think about something difficult or unpleasant. It seems plausible to say that, in such a case, you are aware of a force or power – that you have, by introspection, an “impression” of power. Hume’s only reason for denying this is that you do not understand how this power works, in the sense of knowing, apart from any previous experience, that the effect will occur. But this seems irrelevant: it does not show that you do not experience a feeling of power, force, or connection between your will and its effect. To say that it does show this would be to say that unless the power or force was actually “successful” in bringing about the effect, it did not even exist. But even if you *fail* to hold up the weight or to hold your attention on the unpleasant thought, do you not experience some feeling of (insufficient) force or power? And, if so, why isn’t this an “impression” from which the idea of power can be derived?

Although this objection seems correct, it is not very damaging to Hume’s basic position. For suppose that Hume simply conceded that we have an idea of power derived from an impression of power that we have when we exercise our wills. Could this idea be what we have in mind when we assert, for example, that one billiard ball exerts power or force on another, or that there is a necessary connection between the collision and the movement of billiard balls? Surely not. For a billiard ball is an unconscious, inanimate object; it cannot have an impression of power like the one we may have in voluntary action or deliberate thinking. The same is true of most of the objects that enter into causal relations: they are inanimate, material objects. Clearly, it would be wrong to model our understanding of a necessary connection between events involving these inanimate objects on human volition. But most of the causal relations we *want to understand* involve just such objects: causal relations between the will and its effects are only a tiny sub-class of all causal relations. So, with the possible exception of that sub-class, an idea of necessary connection derived from human volition cannot advance our understanding of causality. Furthermore, even if we do have some idea of a necessary connection between the will and its effects, this idea is not one that supports a deductive inference from cause to effect or from effect to cause.

It is worth noticing that, in at least one place, Hume does seem to concede that we have this limited idea of necessary connection, and to also point out that this does not seriously affect his overall position. He writes:

It may be pretended, that the resistance which we meet with in bodies, obliging us frequently to exert our force, and call up all our power, this gives the idea of force and power. It is this *nisus* or strong endeavour, of which we are conscious, that is

the original impression from which this idea is copied.

(E:67n; S:44n; F:107)

Since Hume does not go on to deny that we do experience this “*nisus* or strong endeavour,” he seems to be conceding that, in such an action as holding up a heavy weight, we do have some impression of force or power. But, in response, Hume says, among other things, that “We attribute power to . . . inanimate matter, which is not capable of this sentiment” (ibid.). This, of course, is the point we are insisting on: even if we do have an idea of power derived from human volition, this idea cannot help us to understand causation in inanimate objects. Hume then adds the revealing remark:

[T]his sentiment of an endeavour to overcome resistance has no known connexion with any event: What follows it we know by experience; but could not know it *a priori*. It must, however, be confessed, that the animal *nisus*, which we experience, though it can afford no accurate precise idea of power, enters very much into that vulgar, inaccurate idea, which is formed of it.

(ibid.)

This reveals once again Hume's assumption that any idea of power or necessary connection worth taking seriously must be such as to license a deductive inference from one event to another; otherwise the idea is not “accurate” or “precise” but only “vulgar” and “inaccurate.” But the remark also betrays his realization that such an idea is, nevertheless, a genuine idea of power – an idea “which is formed of it.”

Having found no impression from which to derive a suitable idea of necessary connection (one that at least would be applicable to causal relations between inanimate objects), Hume next launches a somewhat digressive polemic against philosophers who would try to base our understanding of necessary connection on the notion of acts of God. (Although he mentions no names, he seems to have in mind the Cartesian Occasionalists, especially Malebranche (1638–1715).)

To summarize this attack: Hume begins by asserting that ordinary people think that they actually see the forces operating in nature, except (only) in extraordinary cases that they cannot explain, like earthquakes, plague, and prodigies. In such cases, they attribute the causality to God's intervention. Philosophers, however, realize that

even in the most familiar events, the energy of the cause is as unintelligible as in the most unusual, and that we only learn by experience the frequent *Conjunction* of objects, without ever being able to comprehend any thing like *Connexion* between them.

(E:70; S:46; F:109)

Accordingly, philosophers go much further than do ordinary folk: they infer that the only real cause is always God. Events in the created world are at best only “occasions,” that is, occurrences on the occasion of which God Himself causes the effects. This is Malebranche’s completely general version of the Occasionalism according to which God causes *all* events and so serves as the causal intermediary between physical and mental events.

Needless to say, Hume rejects this view. Apart from eloquently expressing his scepticism about how we could ever know such a view to be true (E:72; S:47–8; F:111), he points out its uselessness for the purpose at hand: namely, the search for an impression of necessary connection.

We are ignorant . . . of the manner in which bodies operate on each other: Their force or energy is entirely incomprehensible: But are we not equally ignorant of the manner or force by which a mind, even the supreme mind, operates either on itself or on body? Whence, I beseech you, do we acquire any idea of it? We have no sentiment or consciousness of this power in ourselves. We have no idea of the Supreme Being but what we learn from reflection on our own faculties. Were our ignorance, therefore, a good reason for rejecting any thing, we should be led into that principle of denying all energy in the Supreme Being as much in the grossest matter. We surely comprehend as little the operations of one as of the other.

(E:72–3; S:48; F:111–12)

The upshot is that appealing to God’s causality cannot increase our understanding of necessary connection beyond that which was attained by reflecting on human volition (which Hume, here again, declares to be nil, and which we have seen to be at least irrelevant to causation involving inanimate objects).

At the start of Section VII part II, Hume summarizes the results of his impression hunt. So far, these results have been totally negative: so long as we consider individual pairs of events – whether these be events involving the things we perceive outside us, or our own volitions and their effects – we find no (suitable) impression of necessary connection. As Hume puts it: “All events seem entirely loose and separate. One event follows another; but we never can observe any tie between them. They seem *conjoined*, but never *connected*” (E:74; S:49; F:113).

So, Hume declares, it looks as if we may be forced to conclude that the idea of power or necessary connection is simply nonexistent, and that these terms are therefore meaningless:

And as we can have no idea of any thing which never appeared to our outward sense or inward sentiment, the necessary conclusion *seems* to be, that we have no idea of connexion or power at all, and that these words are absolutely without any meaning, when employed either in philosophical reasonings, or common life.

(E:74; S:49; F:113)

At this point – just as it looks as if the search for necessary connection has ended in failure – Hume makes the key move in his entire discussion of causality. He asks: when do we first begin to infer, from the occurrence of one event, that some other event will occur? Do we make any such inference the first time that we perceive an event of any particular sort? Certainly not: as Hume has already argued at length, it is impossible, on perceiving an event that is new in our experience, for us to draw from it any inference to another event; the entire weight of his critique of causal reasoning lies behind this negative point. Rather, Hume says, it is only after we have observed that an event of a certain kind is always, in our experience, followed by an event of another kind, that we begin to infer, upon observing an event of the first kind, that an event of the second kind will follow. To use Hume's own famous term, it is only after we experience a “*constant conjunction*” between events – events of kind A always being followed, in our experience, by events of kind B – that we start to infer an event of kind B from observing an event of kind A. Only then do we begin to call one event the “cause” and the other the “effect;” only then do we begin to think that there is some necessary connection between the events. So, the idea of necessary connection must somehow arise from observing many similar pairs of events rather than any individual pairs: it arises from the experience of constant conjunction.

However, the introduction of constant conjunction seems at first not to help at all. For, obviously, if we do not observe a necessary connection between two events of a certain kind, then we do not observe such a connection merely by perceiving many more exactly similar pairs of events. There can be nothing in these similar pairs that was not in the first pair; otherwise they would not be exactly similar. What, then, can possibly be added by observing a multiplicity of similar event-pairs? Only this: a feeling of expectation or anticipation, whenever we perceive an event like the first member of the pair, that an event like the second member will follow. This feeling, which is something in our minds rather than in the events themselves, is the *only* new ingredient added by having the experience of constant conjunction. Therefore, Hume concludes, this feeling IS the impression of necessary connection! This striking thesis is also the point of contact between Hume's theory of causality and his psychological explanation of causal and inductive reasoning. For that explanation, you will recall, appeals to the psychological principle (“habit,” “custom”) that if the mind is presented with repeated cases in which A is followed by B, it will come to expect or anticipate B whenever it perceives A. Hume is now saying that this feeling of anticipation *is* the impression of necessary connection for which he has been searching. In the terminology of the *Treatise*, it is a definite “impression of reflection,” just like joy, pain, or embarrassment.

But if necessary connection is merely a feeling in our minds, why do we think that we have a notion of some necessary connection between events themselves? To answer this question, Hume adds a final but very important point to his account of necessary connection. This is that we *project* our own feeling of expectation or anticipation outward into the observed events, and thereby mistakenly come to think that we are aware of a necessary connection

between the events themselves. As he puts it in the *Treatise* (where this point is made more emphatically than in the *Enquiry*):

[T]he mind has a great propensity to spread itself on external objects. . . . This is the case, when we transfer the determination of the thought to external objects, and suppose any real intelligible connexion betwixt them; that being a quality, which can only belong to the mind that considers them.

(T:167–8)

In the *Enquiry*, he puts the point this way: “[A]s we *feel* a customary connexion between the ideas, we transfer that feeling to the objects; as nothing is more natural than to apply to external bodies every internal sensation, which they occasion” (E:78n; S:52n; F:117). In fact, however, the only relation we observe between the events (“external objects,” “external bodies”) themselves is that of constant conjunction – the fact that events of the same kind as one of them are “constantly conjoined” in our experience with events of the same kind as the other. The upshot, as Hume puts it in the *Treatise*, is:

Upon the whole, necessity is something, that exists in the mind, not in objects; nor is it possible for us ever to form the most distant idea of it, consider'd as a quality in bodies. Either we have no idea of necessity, or necessity is nothing but that determination of the thought to pass from causes to effects and from effects to causes, according to their experienc'd union.

(T:165–6)

Before concluding this part, we should say something about a controversy that has emerged in recent Humean scholarship. Until quite recently, virtually all such scholars have taken Hume to have totally rejected the view that there are necessary connections between events. Against this standard interpretation, however, some writers have recently advanced an interpretation which Kenneth Winkler (1991: 541–79), a contemporary Hume scholar, aptly calls “the New Hume.” According to proponents of this New Hume, Hume denied that humans *know* that there are necessary connections between events, but allowed that such connections may nevertheless exist.¹ Indeed, several of these proponents have gone so far as to assert that Hume himself actually believed that such connections exist, even though he admitted that neither he nor anyone else could know this to be the case.² As evidence for their interpretation, proponents point out that the *Treatise* and the *Enquiry* are rife with passages where Hume uses terms like “secret powers,” “power,” “the ultimate force and efficacy of nature,” “the powers, by which bodies operate,” “secret causes,” “powers and principles on which the influence of . . . objects entirely depends,” and the like, in ways that strongly suggest that he believes that causal powers do exist in objects, even though we cannot know or understand how they operate. The proponents of the New Hume argue also that Hume’s

“noncommittal” sceptical outlook would have prevented him from categorically denying that there are such powers; since this would have been to make a strong claim about reality that he could not really have known to be true (see especially Strawson 1989: 11–12, 94–101, 277).

In response to the New Hume proponents, scholars who defend the standard interpretation of Hume – who defend what Winkler calls “the Old Hume” – have proposed several ways of dealing with the passages in which Hume seems to be referring to “powers,” “secret powers,” “secret causes,” etc. For example, it has been suggested that his use of such expressions is not genuinely referential; that, rather, it involves only “the use, for the sake of argument, of referential expressions favoured by a view that one is opposing” (Bottersill 1990: 205). It has been pointed out that Hume himself, in the footnote in Section IV of the *Enquiry* that we quoted earlier (see p. 77), says that his uses of the word “power” are to be taken in a “loose and popular sense” (cf. Winkler 1991: 545–6, 550). It has been suggested also that Hume’s talk of secret powers and the like may really refer to minute mechanisms and structures rather than to hidden forces (ibid.: 547–50).

The textual issues between proponents of the New Hume and their critics are complex ones that could be settled only by a careful examination of numerous passages and their contexts – something we cannot undertake here. Rather, I shall emphasize only one point that seems to me to be crucial. This is that Hume’s negative remarks about necessary connections between events and powers in objects, in the three places where he most directly and explicitly addresses this topic (Section VII of the *Enquiry*; Book I Part III Section XIV of the *Treatise*; and the *Abstract*), stem directly from his empiricist test for meaning – a test which Hume remained unwilling to give up even when it led him to consequences concerning the self that he found unacceptable. Thus, when Hume makes his negative remarks about necessary connection, he is not just saying that it is *false* that there are necessary connections between events, or that we have no way of knowing whether there are such connections. He is saying something much more radical: namely, that we do not even understand what such a necessary connection would be; so that, if we affirmed that there are such connections, we simply would not understand what we were saying. As Hume puts it in the *Abstract* of the *Treatise*:

Upon the whole, then, either we have no idea at all of force and energy, and these words are altogether insignificant, or they can mean nothing but that determination of the thought, acquired by habit, to pass from the cause to its usual effect.

(T:657; S:134; F:38)

And, again, in the *Enquiry*:

[We] . . . cannot attain any . . . definition, which may point out that circumstance in the cause, which gives it a connexion with its effect. We have no idea of this

connexion, nor even any distinct notion of what it is we desire to know, when we endeavour a conception of it.

(E:77; S:51; F:116)

And this constancy [constant conjunction] forms the very essence of necessity, nor have we any other idea of it.

(E:96*n*; S:64*n*; F:132)

In these passages (and they are not the only ones of this kind: see e.g. T:165–6, quoted on p. 108; and T:168), Hume is saying that we simply do not have any idea or concept of a necessary connection between events. But if this is so, then contrary to what the proponents of the New Hume say, we cannot even really *believe* that there are such connections.³ The closest we can come to such a belief is, as Hume says, to strongly expect that an effect will follow a cause, and to project onto or read into the events themselves this feeling of expectation. But this falls short of genuinely believing that there are necessary connections between the events themselves; for, as Winkler says, “we in no way *refer* to causation as it exists in objects (though we do refer to the objects) when we spread our internal impression of determination onto them” (p. 573).

However, it will be objected, if we cannot even believe that there are necessary connections between events, then neither can we meaningfully *deny* that there are such connections. So what does Hume's rejection of such connections amount to? Winkler offers a plausible answer to this question: he suggests (p. 560; see also pp. 543, 567) that Hume's position consists, not in denying that such connections exist, but rather “in a *refusal to affirm* the existence of [necessary connection between events], a refusal rooted in the belief that there is no notion of [necessary connection between events] to be affirmed or denied, or even entertained as a possibility.”⁴ As Winkler says a little later (p. 576): “Hume needn't say that there is no such thing as objective connection; it is enough for him to say that we cannot *in any way* conceive of it, and that as a result we cannot believe in it.”

It is in this sense that I propose we understand Hume's rejection of necessary connection between events. For not only does this way of understanding Hume fit with the fact that his attack on the notion of necessary connection is deeply rooted in his meaning-empiricism, but it also harmonizes with his attempt to define the causal relation. For, as we shall see in the following part, the notion of a necessary connection between events has no place at all in Hume's definition of causality.

3 Hume's analysis of causation

Immediately after presenting his account of necessary connection, Hume offers us a definition of causation. Indeed, he offers two different definitions – a fact that has caused controversy as to how he should be interpreted. The first definition goes as follows:

Similar objects are always conjoined with similar. Of this we have experience. Suitably to this experience, therefore, we may define a cause to be *an object, followed by another; and where all the objects similar to the first are followed by objects similar to the second.*

(E:76; S:51; F:115)

The second definition goes this way:

The appearance of a cause always conveys the mind, by a customary transition, to the idea of the effect. Of this also we have experience. We may, therefore, suitably to this experience, form another definition of cause, and call it, *an object followed by another; and whose appearance always conveys the thought to that other.*

(E:77; S:51; F:116)

Hume illustrates both definitions with the same example:

We say, for instance, that the vibration of this string is the cause of this particular sound. But what do we mean by that affirmation? We either mean, *that this vibration is followed by this sound, and that all similar vibrations have been followed by similar sounds:* Or, *that this vibration is followed by this sound, and that, upon the appearance of the one the mind anticipates the senses, and forms immediately an idea of the other.* We may consider the relation of cause and effect in either of these two lights; but beyond these, we have no idea of it.

(E:77; S:51–2; F:116)

As scholars have pointed out, Hume's formulations are somewhat loose and ambiguous. One ambiguity is that Hume shifts back and forth between speaking of causes and effects as "objects" and as "events." In both definitions he speaks of them as "objects." But in the surrounding text (especially in the first three paragraphs of part II, where he makes his key claims about necessary connection), he repeatedly refers to them as "events." For example:

All events seem entirely loose and separate. One event follows another; but we never can observe any tie between them.

(E:74; S:49; F:113)

But when one particular species of events has always, in all instances, been conjoined with another, we make no longer any scruple of foretelling one upon the

appearance of the other.

(E:74-5; S:50; F:114)

It appears, then, that this idea of a necessary connexion among events arises from a number of similar instances which occur of the constant conjunction of these events . . .

(E:75; S:50; F:114)

[A]fter a repetition of similar instances, the mind is carried by habit, upon the appearance of one event, to expect its usual attendant, and to believe that it will exist.

(E:75; S:50; F:114)

Altogether, the term “event(s)” occurs twelve times in the first three paragraphs of part II. This suggests that Hume’s own position is that causes and effects are events rather than objects. On the other hand, the term “object” occurs five times in those paragraphs, and Hume does use it in formulating his two definitions. So, the matter cannot be settled on textual grounds alone. We have already seen (Chapter 3, section 2), however, that it is more accurate to regard causes and effects as events than as objects. We can speak of the objects themselves as causes only in a derivative sense, based on the consideration that events are generally changes in objects. For example, we can say that one billiard ball – a certain object – is “the cause” of motion in another, insofar as it is that ball which, by hitting the other, causes the motion. It is obvious, nevertheless, that the cause is not just the ball as such, but its collision with the other ball, which is an event. It is even more obvious that the effect is not the other ball itself, but rather that ball’s movement following the collision, which is also an event. In what follows, therefore, we shall interpret Hume’s definitions as pertaining to events rather than objects.

A second ambiguity arises from the generalization contained in the first definition. This generalization – that “all the objects [events] similar to the first, are followed by objects [events] similar to the second” – presumably refers to *all* such event-pairs, past, present, and future. But when Hume illustrates the definition with the example of the vibration and the sound, he formulates the corresponding generalization in the past tense: all similar vibrations, he says, “*have been followed by similar sounds.*” This refers only to past vibration–sound conjunctions. Now we have seen that for Hume, there is a major epistemological problem about how we can *know* that events which have been constantly conjoined in the past will be constantly conjoined in the future: this is none other than the problem of induction. Perhaps, then, it is Hume’s scepticism about the possibility of justifying induction that lies behind his shift to the past tense in the example.

Nevertheless, we shall interpret the generalization contained in Hume’s first definition as covering all cases – past, present, and future. There are two reasons for doing so. The first is

a textual one. Hume states his first definition of causation on three other occasions – twice in the *Treatise* (T:170, 172) and once in the *Abstract* (T:649–50; S:129; F:33). It is true that on these occasions, he complicates matters somewhat. Instead of saying simply that the cause is “followed by” the effect, he breaks this statement down into two points: the cause is (a) *spatially contiguous* with and (b) *temporally prior* to the effect. This two-point interpretation of “followed by,” which Hume presents in the *Treatise* but drops in the *Enquiry*, is problematic. For some causes, such as magnets attracting iron filings and heavenly bodies exercising gravitational “pull,” at least seem to act at a distance: they are not spatially contiguous with their characteristic effects.⁵ Further, some causes are simultaneous with their effects (e.g. stepping in sand and a footprint’s formation) and thus are not temporally prior to their effects. Indeed, a few contemporary philosophers have even claimed that an effect could conceivably occur *before* its cause! We shall not pursue these points here (though simultaneity of cause and effect will come up again in part 4 in relation to causes as sufficient and necessary conditions). The point we wish to make, rather, is that on each of the other three occasions when Hume gives his first definition of causation (as well as in the *Enquiry* itself), he formulates the generalization that it contains as an absolute universal – one that covers all past, present, and future cases. Thus, it seems safe to assume that he does not intend to restrict the generalization only to past cases.

The other reason for interpreting the generalization as universal is philosophical rather than textual. One of the most important and influential ideas in Hume’s theory of causality is that every particular cause-and-effect statement is implicitly a generalization. The statement that Louis XVI’s decapitation *caused* his death, for example, is true only because decapitating a human being is *always* followed by death. If decapitating a human being were sometimes but not always followed by death, then it would not be true that decapitating Louis XVI *caused* his death. Rather, we would have to look for some other factor in the situation – one which *is* always followed by death – and say that *it* was the true cause of Louis’s death. The statement that the decapitation did cause his death is true *because* the decapitation– death generalization is universal and without exception: it is a *law of nature*. Hume’s idea, then, can also be put this way: every particular causal statement involves at least one law of nature. Now laws of nature do not hold just for past cases; they hold for all cases. Of course, if the problem of induction is taken to imply inductive scepticism, then it calls into question whether we can know that there are such laws. That is why inductive scepticism cuts deep: it concerns the very possibility of science, which seeks to discover laws of nature or scientific laws. The fact remains, however, that laws of nature are *supposed* to hold for all cases; otherwise they would not be laws. This is the fundamental reason for interpreting the generalization in Hume’s first definition as covering not only past cases, but all present and future cases as well.

In line with the two interpretive points we have made – that both definitions pertain to events rather than objects, and that the generalization in the first definition pertains to all cases rather than only to past cases – we may formulate Hume’s two definitions as follows, letting “E₁” and “E₂” stand for any two particular events that are related as cause and effect:

DI: E_1 causes $E_2 = df$

(1) E_1 is followed by E_2 ;

and

(2) all events similar to E_1 are followed by events similar to E_2 .

D2: E_1 causes $E_2 = df$

(1) E_1 is followed by E_2 ;

and

(2a) observing an event similar to E_1 always leads us to expect an event similar to E_2 .

The fact that Hume gives variant definitions of causation obviously calls for comment. Notice, first of all, that the definitions are different: a pair of events could satisfy *DI* without satisfying *D2*, and conversely a pair of events could satisfy *D2* without satisfying *DI*. Suppose for example that while events similar to E_1 are macroscopic ones that we frequently observe, events similar to E_2 are microscopic ones that science has not yet discovered. Then it could be true that events like E_1 are always followed by events like E_2 , but false that events like E_1 always (or ever) lead us to expect events like E_2 . In such a case, *DI*(2) would be satisfied, but *D2* (2a) would not be satisfied. Or suppose that we had been conditioned to expect rain after every solar eclipse, because, as it happened, each solar eclipse we knew of had been followed by rain. It could nonetheless be the case – in fact you and I know it is the case – that not every solar eclipse is followed by rain. So in such a situation, *D2*(2a) would be satisfied, but *DI*(2) would not. It is clear, then, that *DI* and *D2* are not equivalent. It is worth noting, however, that there is a relationship between them. The contemporary philosopher Barry Stroud, in his book *Hume*, helpfully describes this relationship as follows:

The relation between them is something like this. Any events or objects observed to fulfil the conditions of the first “definition” are such that they will fulfil the conditions of the second “definition” also. That is to say that an observed constant conjunction between As and Bs establishes a “union in the imagination” such that the thought of an A naturally leads the mind to the thought of a B. That is just a fundamental, but contingent, principle of the human mind.

(Stroud 1977: 90)

Otherwise put, Stroud's point is that any events that have been *observed* not to violate *DI*(2) will, because of the principle of association Hume has put forward, satisfy *D2*(2a).⁶

As previously mentioned, the nonequivalence of Hume's two definitions has led to controversy concerning his intentions. Some scholars have argued that only Hume's first definition represents his real view (see Robinson 1962). Some have argued that he has two different theories that can be integrated as one (Beauchamp and Rosenberg 1981). Stroud

(1977: 89) holds that Hume never intended, strictly speaking, to give a definition of causation. Recently, Don Garrett (1997: 107–17) has argued that Hume's two definitions can be interpreted in such a way that they turn out to be equivalent, and that Hume accepts both of them. We shall not delve into this debate. Rather, let us look into the implications of Hume's definitions for their own sake, without worrying further about his exact intentions. Specifically, let us ask this question: which of Hume's two definitions is a definition of causation as it occurs objectively in nature, regardless of whether there are any people or observers or sentient beings? The answer is obvious: only *D1*. For *D2* refers to observations, expectations, and the triggering of the latter by the former. But surely cause-and-effect relationships occur in nature independently of these things, and could have occurred even if there had never been any sentient beings capable of making observations or having expectations. To quote Stroud once again:

[T]hings could fulfil the conditions of the first "definition" even if there were no minds at all, or if minds were very different from the way they actually are. The existence and precise nature of minds is irrelevant to the question whether members of one class of things are regularly followed by members of another class.

(Stroud 1977: 90)

Let us take it, then, that *D1* is supposed to be the complete definition of causation, as it exists objectively in nature. *D2*, by contrast, makes only a contingent claim about *observed* cases of causation, as Stroud suggests. What are the implications? One of them, as we have already seen, is that every particular causal statement – nature linking events like E_1 and events like E_2 . But there is another, and striking, every statement of the form " E_1 causes E_2 " – implies that there is some law of implication. Notice that *D1* makes absolutely no reference to necessary connection or to necessity. The sole reference to necessary connection occurs, under another name, in *D2*; since necessary connection just *is* the feeling of expectation mentioned in *D2*. *D1*, by contrast, does not involve necessary connection. Instead, it involves only "constant conjunction" – one type of event being always accompanied or followed by another type of event. The essence of causation, then, is not necessary connection (which exists only in our minds), but constant conjunction. To put it differently, the essence of causation is not necessity, but *regular succession* or *regularity*.

An example should help to make this implication of *D1* more vivid. Think of a situation in which automobiles are being tested by the manufacturer for crash resistance. By signals from a radio transmitter, test cars are driven into a brick wall at various speeds, to determine the extent of damage that results. Imagine that a particular car is driven into the wall at 5 miles per hour. Imagine that upon impact the front bumper crumples. We may say, then, that *the impact of the car against the brick wall was followed by the crumpling of the car's bumper*. In other words, clause 1 of *D1* is satisfied. Now, what are we *adding* when we say that the impact *caused* the crumpling? According to *D1*, the only thing we are adding is that *all*

impacts of this kind are followed by crumplings of that kind – a pure statement of regularity. We are *not* adding that the impact *necessitated* the crumpling, or that there was some necessary connection between the two events. If we do say this, that is because we are projecting our own anticipation or expectation of the bumper's crumpling onto the events themselves – a mistake, as Hume has argued.

Finally, *DI* also has implications concerning the character of laws of nature. For, as we noted above, clause 2 of *DI* – the generalization that all events similar to E_1 are followed by events similar to E_2 – states the law of nature involved. But this generalization itself makes no reference to necessity: it is simply a universal exceptionless generalization. Thus, *DI* implies that laws of nature are only exceptionless but contingent regularities between kinds of events. They do not embody any element of necessity.

4 The regularity theory: some objections and replies

The analysis of causation given in Hume's first definition, and the view of laws of nature it embodies, are now commonly called the "regularity theory" of causality. There is a large and complex literature, both "pro" and "con," on the regularity theory. Indeed, this literature is only a part of a larger literature on a group of interrelated topics, including not only causation itself but also the nature of natural laws, the nature of events, scientific explanation, and others. We cannot possibly survey this literature here, nor even touch on all of the issues that it raises. Rather, we shall try only to bring out what makes the regularity theory appealing, and to defend it against a number of objections.

The main appeal of the regularity theory is probably this: it takes some of the mystery out of causality. For the notion of a necessary connection between events, of necessity *in nature* or *natural necessity*, is a puzzling and mysterious one. Hume's attempt to pinpoint something observable that would answer to this notion, whatever else may be said about it, serves at least to bring out the elusive character of such necessity. Compare it with logical necessity – with the necessity by which philosophical questions about the nature of logical necessity, there are several ways the conclusion q follows from the premisses *if p, then q* and p . Although there are of getting a grasp of the notion. We can point out that it would be a contradiction to assert those premisses and deny that conclusion. We can point out that there is no possible world in which both premisses, p and *if p, then q*, are true and the conclusion, q , is false. We can point out that the statement, "*{[if p, then q] and p}, then q*," is a truth of logic or a "tautology" – a statement that remains true no matter what combination of truth-values we assign to p and q individually. But none of these points carries over to causal relationships or laws of nature. It is never a contradiction, as Hume showed, to affirm that one event occurred but that another – distinct – event did not. There are possible worlds in which events that are conjoined in our world are disjoined, and in which different laws of nature obtain. The laws of nature are not truths of logic or tautologies. Thus, we cannot understand natural necessity in terms of logical necessity. When this (essentially Humean) insight is combined with Hume's point that nothing we can observe

answers to the name of natural necessity, that notion seems mysterious and incomprehensible. By contrast, the notion of regularity seems clear, crisp, and unmysterious. We can observe regularities obtaining between events. We can use refined scientific methods to find the real regularities in nature, and to distinguish them from those suggested by superficial examination. Putting aside Hume's worries about induction, we can use those regularities to predict future events and to explain past events. So if causation and laws of nature are simply matters of regularity, our world is a little less mysterious than it would be if they embody some sort of non-logical, natural necessity.

But while the regularity theory is very appealing, it also faces difficult objections. In the subsections that follow, we shall discuss four of them.

The problem of similarity

In a number of writings, the contemporary American philosopher Richard Taylor attacks the regularity theory (e.g. 1967: 56–66; 1992: Chap. 10). In *Metaphysics* (1992); Taylor first offers the following test for the correctness of *DI*: in order for that definition to be correct, it must not be possible to describe any case where its *analysandum* or *definiendum* (the expression to be analyzed or defined, namely, “ E_1 causes E_2 ”) is true while its *analysans* or *definiens* (the defining expressions, namely, clauses 1 and 2 of *DI*) are false, or vice-versa. Taylor then argues that *DI* either fails the test or is circular. He focuses in on the question: what does “similar” mean in the definition? If it means “exactly similar” (i.e. similar in every respect, presumably including even spatial and temporal positioning), then the only event which is similar to E_1 is E_1 itself, and the only event which is similar to E_2 is E_2 itself (Taylor 1992: 92). But then if clause 1 of *DI* is satisfied – if E_1 is followed by E_2 – clause 2 is automatically satisfied as well. For since “similar” means “exactly similar,” no other event can be similar to E_1 , and likewise no other event can be similar to E_2 : so, of course, all events similar to E_1 are followed by events similar to E_2 ! But this means that we can easily describe cases where both clauses 1 and 2 of *DI* are true while “ E_1 causes E_2 ” is false (i.e. where the *analysans* is true but the *analysandum* is false). Any case where E_1 is followed by E_2 , but E_1 does not cause E_2 , will do. For example: I look at a red traffic light, and it turns green. The only event exactly similar to my act of looking at a red light is that very act itself, and the only event exactly similar to the event of the light turning green is that very event itself. So – assuming still that “similar” means “exactly similar” – all events similar to that act of looking are followed by events similar to that event of turning green. But it would be ridiculous to say that my looking at the red light causes it to turn green.

On the other hand, suppose that “similar” means “similar to a high degree.” Then we can easily describe cases where the *analysandum* is true but the *analysans* is false. For example, suppose that striking a certain match (E_1) causes it to ignite (E_2). Then although clause 1 is satisfied E_1 – is followed by E_2 – clause 2 is not satisfied. For many events that are similar to a high degree to E_1 , such as striking a damp match, or a dry match on too smooth a surface, are not followed by events similar in a high degree to E_2 (Taylor 1992: 92). So now we have

a case where "E₁ causes E₂" is true, but where, due to the falsity of clause 2, the conjunction of clauses 1 and 2 is false – i.e. it is a case where the *analysandum* is true but the *analysans* is false.

It might be thought that the problem can be solved by letting "similar" mean "exactly similar except for space–time location." But on that interpretation of "similar," the match-striking case still shows that the *analysandum* might be true while the *analysans* was false. For a small variation occurring after an event "similar" to E₁ (striking the match), such as a light breeze arising, might result in the ignition or flaming up being a bit different from E₂ (the match's ignition); so that clause 2 of the *analysans* would be false: we would have a case where an event "exactly similar" to E₁ was *not* followed by an event "exactly similar" to E₂. Yet it would still be true that E₁ causes E₂.

Finally, suppose that we try to solve the difficulty by letting "similar" mean "similar in relevant respects." Then the analysis is spoiled; for what does "relevant respects" mean? Well, it can only mean "causally relevant respects" (Taylor 1992: 92). But this turns *DI* into a circular definition, for clause 2 now means:

- (2b) all events that in causally relevant respects are similar to E₁ are followed by events that in causally relevant respects are similar to E₂.

But this is, basically, to define causation in terms of itself. The upshot, then, is that *DI* is either incorrect or circular.

It must be admitted that Taylor's objection to *DI* is a powerful one. Nevertheless, the objection is not fatal to the regularity theory; for the objection hinges on the fact that Hume formulates the regularity – in clause 2 of *DI* – in terms of *similarity* or resemblance. But there is another way to formulate this regularity: namely, in terms of the kinds, classes, or species of event involved. In a passage quoted earlier, we saw that Barry Stroud does not hesitate to put Hume's view in this manner; he speaks of "members of one class of things [being] regularly followed by members of another class" (1977: 90). Indeed, Hume himself sometimes speaks in this way, as for example when he says:

But when one particular *species of events* has always, in all instances, been conjoined with another, we make no longer any scruple of foretelling one upon the appearance of the other . . .

(E:74–5; S:50; F:114; emphasis added)

Let us try any other balls *of the same kind* in a like situation, and we shall always find that the impulse of the one produces motion in the other.

(T *Abstract*:649; S:129; F:33; emphasis added)

Now if the generalization in *DI* is reformulated in terms of kinds or species or classes of event, then it must say that all events of one kind (species, class) are followed by events of another kind. Furthermore, *DI* as a whole must be modified, in order to bring it into line with this reformulation. The modified definition would go as follows:

D3: E_1 causes $E_2 = df$:

- (1) E_1 is followed by E_2 ;
- (2) all events of kind K_1 are followed by events of kind K_2 ;
- (2) E_1 is of kind K_1 and E_2 is of kind K_2 .

This definition may, of course, be vulnerable to other objections. But one thing seems clear: it is not vulnerable to Taylor's objection. For that objection hinges on the troublesome notion of similarity or resemblance – a notion that does not appear in *D3*.

There is a reason why Hume himself speaks mainly in terms of similarity between events rather than kinds of event: he is a nominalist. Nominalism is a theory about what it is for two or more things to have a common property. Roughly speaking, the theory says that things have a common property when they are similar – when they resemble each other. Furthermore, since every property defines a class or kind, nominalism is also a theory about what it is for two or more things to belong to the same class or to be of the same kind: again, the things are similar, or resemble each other. Now when Hume puts forward his regularity definition of causation, he does so within the framework of his nominalism. That is why he expresses the definition in terms of similarity. But in assessing the regularity theory, it seems only fair to separate it from the issue of nominalism. One question is whether it is possible to formulate a satisfactory definition of causation *purely* in terms of regular succession between events of certain kinds or species, without appealing to the notion of necessary connection. If the answer is yes, then the regularity theory is vindicated. Another question is whether nominalism is justifiable. If the answer is yes, then there seems no reason in principle why all statements referring to kinds or species of event could not be reformulated as statements referring only to resembling events – in which case Hume's combined regularity view of causation and nominalism would be vindicated too. But in assessing the regularity theory as such, it is necessary to separate it from the question of nominalism.

The trouble with Taylor's objection is that it depends precisely on not separating these issues. Taylor attacks the regularity theory by bringing out difficulties that stem only from Hume's nominalistic formulation of it. To see this, notice that nominalism itself could be attacked in much the same way as Taylor attacks *DI*. The nominalist holds that two things have a common property provided that they are similar. Now, Taylor could say: what does it mean for them to be similar? If it means that they are exactly similar, then they are not really two things but one. If it means that they are similar to a high degree, they might nonetheless not share a common property. If it means that they are similar in relevant respects, then the theory is circular, because those "respects" can signify only the common

property in question. To answer this objection, a nominalist would have to refine the theory in some way. Whether or not this can be successfully done is an interesting and controversial question in its own right. But this question is quite independent of whether causation can be defined in terms of regularity.

Causal laws versus accidental generalizations

Without doubt, the most common objection to the regularity theory is that it is incapable of distinguishing between two importantly different things: genuine causal relationships, and regular but non-causal relationships. As applied to *D3*, the objection is that we can easily describe cases where all three clauses of the *analysans* are satisfied, but E_1 does not cause E_2 . For example, suppose that whenever the noon schoolbells ring in Washington, DC, students in New York City schools go to lunch. Suppose that Jefferson High is a school in Washington, DC, and that Dewey High is a school in New York City. On a given day, the noon schoolbell rings at Jefferson High, and the students at Dewey High troop off to lunch. Then it is true (i) that the bell's ringing at Jefferson High is followed by students going to lunch at Dewey High, (ii) that all events of the kind: noon schoolbells ringing in Washington, DC are followed by events of the kind: students going to lunch in New York City schools, and (iii) that the bell ringing at Jefferson High is an event of the kind: noon schoolbells ringing in Washington, DC, and students going to lunch at Dewey High is an event of the kind: students going to lunch in New York City schools. Yet it is clearly false that the bell ringing at Jefferson High causes students to go to lunch at Dewey High. Therefore, clauses 1, 2, and 3 of *D3*'s *analysans* are all satisfied, but E_1 does not cause E_2 . So, *D3* must be incorrect.⁷

A regularity theorist would reply that the reason why E_1 does not cause E_2 in this example is quite obvious: the generalization stated in (ii) – that whenever noon schoolbells ring in Washington, DC, students in New York City schools go to lunch – is not a causal law or a law of nature. Rather, it is a true but accidental or coincidental generalization. In order for clause 2 of *D3*'s *analysans* to be satisfied, (ii) would have to cite a causal law, rather than just an accidental generalization.

This reply, however, leads directly to the question: how does the regularity theorist propose to distinguish between a causal law and an accidental generalization? This is a crucial question for the regularity theorist. For unless an answer is forthcoming, there is no way to save *D3* from the objection that its *analysandum* can be false while its *analysans* is true. As the contemporary British philosopher J. L. Mackie (1917–81), put the matter in his highly regarded *The Cement of the Universe*: “The problem . . . of distinguishing causal from accidental regularities, is the great difficulty for any regularity theory of causation” (1974: 196).

One way to make the distinction, of course, would be to say that a causal regularity always involves a necessary connection, while an accidental regularity does not. But this is precisely what a regularity theorist cannot say, without giving up the entire theory. The

problem such a theorist faces, then, is to distinguish between a causal law and an accidental generalization *without* appealing to the notion of a necessary connection between a cause and its effect. To deepen our understanding of this problem, let us compare two statements: (a) Whenever a human being is beheaded, he or she dies. (b) Whenever John has eggs for breakfast, Mary has toast. We know, of course, that (a) is true, for it involves a law of nature. Suppose that (b) is also true. In other words, suppose it just happens to be true, throughout John and Mary's entire lives, that each time John eats eggs at breakfast, Mary eats toast. Both (a) and (b), then, are true and exceptionless generalizations. Yet, it is obvious that while (a) exemplifies a causal law, (b) is an accidental generalization. What is the difference between them?

One difference to which philosophers have called attention is this. Suppose that John is, happily enough, not beheaded. Then we can still be sure that (c) If John had been beheaded, he would have died. This statement is of a type called a *counterfactual conditional* or *contrary-to-fact conditional*. Such a statement says that if something which did not happen – something which is, so to speak, counter or contrary to the facts – *had* happened, then something else would have happened. Now an important feature of causal laws is that they allow us to infer various counterfactual conditionals. As it is usually put, causal laws “sustain” counterfactual conditionals. In the case at hand, (a) allows us to infer (c): it sustains (c). By contrast, a merely accidental generalization like (b) does not sustain any counterfactual conditional. To see this, suppose that on the morning of July 14, 2000 John does not have eggs for breakfast. Then we *cannot* infer from (b) that (d) If John had had eggs for breakfast on July 14, 2000 then Mary would have had toast. In other words, the fact that all occasions on which John has eggs for breakfast happen to be occasions on which Mary has toast does not enable us to infer that, if John had had eggs on a certain morning when he in fact did not, Mary would have had toast. On the other hand, the causal law that all occasions on which some person is beheaded are occasions on which that person dies does enable us to infer that, if John, who in fact was not beheaded, had been beheaded, then he would have died.

Now some philosophers argue that this difference shows that the problem faced by the regularity theory – that of distinguishing between causal laws and accidental generalizations without appealing to natural necessity – is insuperable. For it shows, they think, that (a) possesses a special necessity which (b) does not.

A regularity theorist, however, need not concede this point. For the theorist can point to a difference other than necessity between (a) and (b), by virtue of which (a) is a causal law but (b) is not. This is that (b), unlike (a), is not sufficiently general to sustain counterfactuals. For (b) refers to two *particular individuals*, John and Mary, and says that every occasion on which John has eggs for breakfast is an occasion on which Mary has toast. This provides no basis for saying that if, on a given occasion when John did not have eggs for breakfast, he had eaten eggs for breakfast, then Mary would have had toast. By contrast, (a) pertains to two *kinds* of event, and says that every event of kind K1 (beheading of a human) is followed by

an event of kind K2 (death of that human). This *does* provide a basis for saying that if, on a given occasion when an event of kind K1 did not occur, an event of kind K1 had occurred, then an event of kind K2 would have occurred. It would seem, then, that what prevents (b) from sustaining counterfactuals, and so disqualifies it from being a causal law, is its reference to particular individuals. This harmonizes with something that we all know about laws of nature: laws of nature are not restricted to particular individuals. They pertain to kinds or types of event, regardless of particular circumstances of place and time. This is also why it cannot be a causal law that whenever noon schoolbells ring in Washington, DC, schools, students go to lunch in New York City schools. This generalization, in virtue of its reference to specific cities, fails to sustain counterfactuals, and is therefore disqualified from being a law of nature.

Although the requirement that a causal law cannot refer to particular individuals seems to dispose of the Washington–New York City case and the John–Mary case, there are examples of greater difficulty. Consider the following case, given by Richard Taylor (1992: 93–4). Suppose that we have a batch of 1,000 matches. Suppose that we decorate each of the matches with a special design, D, which is, say, a pattern of alternating red and green lines, around its stem. Suppose that in the entire history of the universe, no other match will ever be decorated with D. Finally, suppose that all of our D-decorated matches are struck against a piece of fresh sandpaper, and that all ignite. Then it is true that all strikings of D-decorated matches are followed by ignitions. Yet, this generalization is no law of nature; for it does not sustain the counterfactual claim that if some match, not from the original batch of 1,000, and so not decorated with D, had been decorated with D and struck against a fresh piece of sandpaper, it would have ignited. But why does it not sustain this counterfactual? The only reason that can be given, according to Taylor (1992: 94), is that there is no necessary connection between a match's having D and its igniting when struck, as there is, for instance, between its having a certain chemical composition and its igniting when struck.

In order to answer the objection, a regularity theorist must, again, be able to disqualify Taylor's generalization as a causal law without appealing to the notion of necessary connection. It may be tempting to suppose that this could be done in much the same way as before, by saying that the generalization does, after all, refer to certain specific matches. The fact that there are 1,000 of them, or that they do not have names like "John" or "New York City," does not prevent the generalization from referring illegitimately to certain individual things. Taylor, however, could reply that his reference to the 1,000 matches was made only in order to help us understand the example; no such reference is actually contained in the generalization at issue. For that generalization simply says that all D-decorated matches ignite when struck against fresh sandpaper. Unlike the Washington–New York City and John–Mary generalizations, it does not mention any individual things – it does not refer to the particular matches that happen to fall under it. Putting it in the terminology of *D3*, Taylor's generalization says only that all K1 events – events of the kind: striking of a D-decorated match against fresh sandpaper – are followed by K2 events – events of the kind: ignition of a D-decorated match. Since each of the 1,000 D-decorated matches that will ever exist does ignite when so

struck, this generalization is perfectly true or exceptionless. Yet it is no law of nature, and the regularity theorist cannot tell us why.

The regularity theorist need not surrender. For ask yourself: would we really assert, in the case Taylor describes, that all K1 events are followed by K2 events? No. For to be in a position to assert that all K1 events are followed by K2 events, more is required than merely passive observation of a regularity between such events. We have to vary, in a controlled way, the conditions under which the events occur, to see whether K1 events are still followed by K2 events. As Kant put it:

When Galileo caused balls, the weights of which he had himself previously determined, to roll down an inclined plane; when Toricelli made the air carry a weight which he had calculated beforehand to be equal to that of a definite volume of water . . . or when Stahl changed metals into oxides, and oxides back into metal, by withdrawing something and then restoring it, a light broke upon all students of nature. They learned that reason has insight only into that which it produces after a plan of its own, and that it must not allow itself to be kept, as it were, in nature's leading-strings, but must itself show the way with principles of judgment based upon fixed laws, constraining nature to give answer to questions of reason's own determining. Accidental observations, made in obedience to no previously thought-out plan, can never be made to yield a necessary law, which alone reason is concerned to discover. Reason, holding in one hand its principles, according to which alone concordant appearances can be admitted as equivalent to laws, and in the other hand the experiment which it has devised in conformity with these principles, must approach nature in order to be taught by it. It must not, however, do so in the character of a pupil who listens to everything that the teacher chooses to say, but of an appointed judge who compels witnesses to answer questions which he has himself formulated.

(Kant 1963: 20)

Apart from the phrase "a necessary law" (to which a regularity theorist would of course object) and the rationalistic-sounding references to "reason," this passage provides an eloquent antidote to Taylor's example. For the scenario his example suggests is the very antithesis of the scientific practices Kant describes: we have a batch of matches that have many properties besides D; we find that they all light when struck; and we conclude, in spite of the fact that we cannot control the variables involved, that it is D, rather than any of the other properties, which is always followed by ignition. Scientific laws are not established in such a manner.

The trouble is that Taylor sets up the case in such a way that no genuine testing of the generalization is possible. By hypothesis, no matches other than those we have ourselves decorated with D will ever, in the history of the universe, have D. This means that we cannot, for example, paint D on a wet match, or on a match that has already been lit, so as to

test the generalization that *all* D-decorated matches ignite when struck on fresh sandpaper. By hypothesis, we cannot perform the tests that any person with even a rudimentary understanding of scientific method knows are indispensable – tests which, in this case, would include putting D on matches not included in the original 1,000, giving them properties different from those of the one thousand (e.g. wetness), and striking them against fresh sandpaper. But since we cannot do this, it seems that instead of asserting that all K1 events are followed by K2 events, all we can reasonably say is: “So far, all K1 events have been followed by K2 events. But, unfortunately, we will never be able to know whether *all* K1 events are followed by K2 events. For all the D-decorated matches that we tested also had many other properties, such as being dry, not previously struck, etc. But since no other matches will ever have D, we will never know whether a match that has D but is wet, or has D but has been previously lit, ignites when struck. So, we will never know whether the regularity we observed between K1 and K2 events is a universal one.” Suppose, on the other hand, that a generalization of the form “all K1 events are followed by K2 events” is scientifically tested and, as a result of such testing, held to be true – to be a law. Is there anything in this situation to indicate that the scientists have discovered something more than just an exceptionless regularity between K1 and K2 events? Must we say, for example, that they have found a *necessary* connection between such events? This seems in no way necessary; nor, as Hume has shown, is it at all clear what “necessary” means in this context.

The fundamental point can be put as follows. Consider the generalization:

- (M) All strikings of D-decorated matches are followed by ignition of Ddecorated matches.

For this generalization to express a law of nature, it must not mean just that

- (M¹) All strikings of D-decorated matches that actually exist are followed by ignition of D-decorated matches.

Rather, it must mean that

- (M²) All strikings of D-decorated matches that actually exist are followed by, and all strikings of D-decorated matches that might have existed would be followed by, ignition of D-decorated matches.

But Taylor has described only a case where (M¹) obtains, not one where (M²) obtains. Further, if we were satisfied that (M²) was true, then we would be justified in regarding (M²) as a law of nature.⁸

Taylor might object that the counterfactual phrase (“that might have existed would be followed by”) imports into (M²) the very notion of necessity that the regularity theory seeks to avoid. But the regularity theorist can reply that this is not at all obvious; for, if that

phrase were replaced with the future-tense phrase “that will exist will be followed by,” then it would not necessarily import the notion of a necessary connection between a cause and its effect. It would only express a prediction that the generalization will hold in the future, and is thus a genuinely exceptionless generalization. By the same token, there seems no reason to think that the counterfactual phrase, which the hypothesis that no other D-decorated matches will ever exist forces us to adopt, must import this notion of necessity.⁹

We conclude that, for all Taylor's example shows, the regularity theorist may be right in holding that the difference between an accidental generalization and a causal law is, at bottom, a matter of generality rather than necessity.

Causes as sufficient conditions and as necessary conditions

As we saw earlier, Hume states his first definition of causation this way: “[W]e may define a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second*” (E:76; S:51; F:115). Immediately after giving this definition, however, Hume adds a very puzzling remark, which I did not include when first quoting the passage, and have simply ignored until now. He says: “Or, in other words, where, *if the first object had not been, the second never had existed*” (E:76; S:51; F:115). The reason this remark is puzzling is that, despite Hume's preface “in other words,” it is not at all equivalent to the definition he has just given; nor is it even implied by that definition. This can be seen by looking at *D3*, which is modeled closely on Hume's own definition. Clauses 2 and 3 of *D3* imply that if E_1 occurs, then E_2 occurs. So, *D3* implies that when E_1 causes E_2 , E_1 is a *sufficient condition* for E_2 . Clauses 2 and 3 do not imply, however, that if E_1 does not occur, then E_2 does not occur. So, *D3* does not imply that when E_1 causes E_2 , E_1 is a *necessary condition* of E_2 . Yet, in his remark Hume seems to be saying that *D3* does carry this implication.

Hume makes this puzzling remark only once; so it might be tempting to dismiss it as just a careless slip. But the remark cannot be treated so lightly; for some contemporary philosophers would say that it betrays Hume's fleeting realization that his definition of causation leaves out something important. In other words, they would say that an adequate analysis of causation *should* imply that a cause is not just a sufficient condition for its effect, but also a necessary condition for its effect.¹⁰ Consider, for example, the statement that striking a match causes it to ignite. On *D3*, this statement implies only that striking the match is sufficient for it to ignite – that if the match is struck, then it ignites. But does not the statement also imply that striking the match is necessary for it to ignite – that if the match is not struck, it will not ignite? Some contemporary philosophers, J. L. Mackie and Richard Taylor for example, would claim that it does. They would argue, furthermore, that although Hume's definition can be expanded so as to carry this implication, the expanded definition runs into serious difficulty. Let us consider how the expanded definition would go, what difficulties result, and how Hume (or a regularity theorist) might deal with them.

In order for *D3* to imply that a cause is a necessary condition for its effect, we would have to add the following statement to clause 2: “and all events of kind K2 are preceded by events of kind K1.” The expanded clause would then read:

- (2') All events of kind K1 are followed by events of kind K2, and all events of kind K2 are preceded by events of kind K1.

This statement, together with clause 3 of *D3*, implies not only that if E_1 occurs, then E_2 occurs, but also that if E_2 occurs, then E_1 occurs. So if 2' is substituted for 2 in *D3*, then *D3* implies that when E_1 causes E_2 , E_1 is not only a sufficient condition for E_2 , but also a necessary condition for E_2 .

However, this expanded version of *D3* runs into the following difficulty. If E_1 is both sufficient and necessary for E_2 , then E_2 is both sufficient and necessary for E_1 . For the former says that (a) if E_1 occurs then E_2 occurs and if E_2 occurs then E_1 occurs, while the latter says that (b) if E_2 occurs then E_1 occurs and if E_1 occurs then E_2 occurs. But (a) and (b) say exactly the same thing; they differ only in word order. The consequence is that the relationship between E_1 and E_2 turns out to be perfectly symmetrical. But this is an absurd consequence. For we all know that the causal relation is an asymmetrical one – that, for example, striking the match causes it to ignite, but its igniting does not cause it to be struck. The upshot is that, when we expand *D3* so as to accommodate the idea that a cause is a necessary as well as a sufficient condition for its effect, we can no longer distinguish between cause and effect!¹¹

One way to deal with the difficulty would be to insist, as Hume does, that the cause must occur *before* the effect in time – that E_1 must be temporally prior to E_2 . This would certainly restore the required asymmetry of the causal relation.

Unfortunately, however, this solution is not very satisfactory; for there are cases where cause and effect occur simultaneously. Kant cites a case where a ball is placed on a cushion and a hollow is formed, and earlier we mentioned the case of making a footprint in the sand; there are other cases as well. No doubt, not *all* causes are simultaneous with their effects; for, if they were, there would be no such thing as a causal chain or series some of whose members occur before the others. But it has to be admitted that some causes occur at the same time as their effects. (As mentioned earlier, some philosophers have even suggested that some causes occur after their effects!) Now the original – unexpanded – version of *D3*, despite the term “followed by,” can be read so as to allow for simultaneous causes and effects, and so as to preserve the asymmetry of causation as well. For “followed by” can be interpreted to mean simply “accompanied by.” In that case, clause 2 will mean that all events of kind K1 are accompanied by events of kind K2 – that a K1 event never occurs without a K2 event occurring. This allows that the K2 event may occur either after, or at the same time as, the K1 event (or even before it!). So, this reading of clause 2 allows for cases where cause and effect occur simultaneously. It also allows for the possibility that a K2 event occurs without a K1 event occurring. For it means only that all K1 events are accompanied by K2 events, which

does not entail that all K2 events are accompanied by K1 events. So it leaves a sense in which K1 events are “followed by” K2 events, but not vice-versa; it preserves the asymmetry of the causal relation. However, if clause 2 is expanded into 2', then interpreting “followed by” as “accompanied by” no longer suffices to preserve causal asymmetry as well as allowing for cases of simultaneity. For 2' now says both that all events of kind K1 are accompanied by events of kind K2 and that all events of kind K2 are accompanied by events of kind K1, thereby losing the asymmetry. (Of course, “preceded by” also must be read as “accompanied by” if 2' is to cover cases of simultaneity.) Thus, trying to accommodate the idea that a cause is a necessary as well as a sufficient condition for its effect while allowing for cases where cause and effect occur simultaneously leaves us unable to distinguish causes from effects.

To deal with this difficulty more satisfactorily, let us begin by noting a potentially misleading feature of *D3* – one that I have so far ignored. As stated, *D3* suggests that causes are much simpler than they actually are. It suggests, for example, that striking a match is always followed by its igniting; so that striking the match is by itself sufficient for its igniting. But, of course, we know that this is not so. Striking a match is followed by its igniting only if certain background conditions obtain. For example, the match must be dry, there must be oxygen present, etc. John Stuart Mill (1806–73), who defended a more refined version of Hume's regularity theory, expressed the point thus:

It is seldom, if ever, between a consequent and a single antecedent that this invariable sequence subsists. It is usually between a consequent and the sum of several antecedents; the concurrence of all of them being requisite to produce, that is, to be certain of being followed by, the consequent.

(Mill 1973: 327)

Of course, we could read “event of kind K1” in *D3* as *including* all of these conditions or “antecedents.” Indeed, there is a sense in which the oxygen and dryness are just as much “causes” of a match's ignition as is its being struck. For various reasons, we commonly select the striking of the match as “the cause” of its lighting, and regard the other parts of the total cause as background conditions. But these reasons have more to do with economy of language and our practical purposes than with the nature of causality. Nevertheless, if we wished to do justice to the distinction between causes and background conditions in *D3*, we could reformulate clause 2 as follows:

(2A) All events of kind K1 in conditions C are followed by events of kind K2.

Together with clause 3 (see page 119), this implies:

(A) If E_1 occurs in conditions C, then E_2 occurs.

In other words, a cause is sufficient for its effect, provided that certain other conditions obtain.

Just as *D3* can misleadingly suggest that striking a match is by itself sufficient for its igniting, so the expanded version of *D3* – in which 2' is substituted for 2 – suggests that striking a match is always necessary for its igniting. But of course, this too is false. For there are many other possible causes of a match igniting: it might be brought in close contact with a burning match; it might be touched by a very hot piece of metal; it might be exposed to superheated air, etc. In general, a given effect can be produced in a variety of different ways. As Mill put it:

It is not true, then, that one effect must be connected with only one cause, or assemblage of conditions; that each phenomenon can be produced only in one way. There are often several independent modes in which the same phenomenon could have originated. One fact may be the consequent in several invariable sequences; it may follow, with equal uniformity, any one of several antecedents, or collections of antecedents. Many causes may produce mechanical motion: many causes may produce some kinds of sensation: many causes may produce death. A given effect may really be produced by a certain cause, and yet be perfectly capable of being produced without it.

(1973: 435)

This important point provides a solution to our difficulty. For it shows that the idea that a cause is a necessary condition for its effect is not wholly accurate. Rather, a cause is necessary for its effect only on the assumption that no other cause of that effect is operative. It is incorrect, for example, to hold that “striking the match caused it to ignite” implies that if the match had not been struck, it would not have ignited. Rather, this statement implies that if the match had not been struck *and nothing else had caused it to ignite*, then it would not have ignited (assuming that it would not have ignited without *some* cause). Now if one wishes to incorporate this idea into *D3*, then (2') will not serve. Instead, one must add:

(2'') All events of kind *K1* are followed by events of kind *K2* and all events of kind *K2* are preceded by events of kind *K1* or *K3* or *K4* or . . . *Kn*.

Here, *K3*, *K4* . . . *Kn* refer to other kinds of event that are also always followed by events of kind *K2*. In terms of our example, they refer to such events as one match being brought into close proximity to a lighted match, or touched by a piece of hot metal, or exposed to superheated air. Now from 2'' and 3, we cannot deduce that if *E₂* occurs, then *E₁* occurs. Rather, we can only deduce that if *E₂* occurs *and no event of kind K3, K4, or . . . Kn occurs*, then *E₁* occurs. Therefore, the definition no longer implies that a cause is both sufficient and

necessary for its effect. Rather, it implies that a cause is sufficient for its effect, and that it is necessary for its effect given the *nonoccurrence* of other conditions that would also have been sufficient for that same effect. The same point holds true if *D3* is formulated so as to incorporate the common-sense distinction between causes and background conditions. To capture this distinction, 2? must be reformulated as:

- (2A'') All events of kind K_1 in conditions C are followed by events of kind K_2 , and all events of kind K_2 are preceded by events of kind K_1 in conditions C , or K_3 in conditions C , or K_4 in conditions C , or . . . K_n in conditions C .

Here, conditions C refer to the background conditions that must always obtain for an event of kind K_2 to occur. In terms of our example, they refer to such things as the presence of oxygen, the dryness of the match, and the like. Now from 2A'' and *D3*, we cannot deduce:

- (B) If E_2 occurs, then E_1 occurs in C .

Rather, we can deduce only:

- (C) If E_2 occurs, and if no event of kind K_3 occurs in C and no event of kind K_4 occurs in C . . . and no event of kind K_n occurs in C , then E_1 occurs in C .

If the definition did imply both (A) and (B), this would mean that the occurrence of E_1 in C was both sufficient and necessary for the occurrence of E_2 . The consequence, again, would be to represent the causal relation as symmetrical. But the combination of (A) and (C) does not have this consequence; it allows for the asymmetry that the causal relation requires.

However, it might now be objected that this asymmetry has been preserved only because the expanded definition is incomplete. To make it complete, one must add still another condition to *D3*, namely:

- (4) E_2 is not preceded by an event of kind K_3 , or K_4 , or . . . K_n .

Once clause 4 is added to clauses 1, 2'', and 3, the asymmetry of the causal relation is once again denied. For 2? and 3 imply that if E_1 occurs, then E_2 occurs; while 2'', 3, and 4 imply that if E_2 occurs, then E_1 occurs. So the complete, expanded version of *D3* implies that E_1 is both sufficient and necessary for E_2 – thus making the relationship between E_1 and E_2 perfectly symmetrical. Furthermore, it does not help to reformulate 4 as:

- (4A) E_2 is not preceded by an event of kind K_3 in C , or an event of kind K_4 in C , or . . . an event of kind K_n in C .

For 2A'', 3, and 4A imply (B), which, together with (A), again makes the relationship between E_1 in C and E_2 symmetrical.

The reply to this objection is that a regularity theorist can reasonably refuse to add 4 or 4A to the definition. For once either of these clauses is added, the definition is no longer an analysis of "E₁ causes E₂." Rather, it is an analysis of "E₁ causes E₂ and nothing else causes E₂," or "E₁ is the (only) cause of E₂." No doubt when we make ordinary causal judgments, we often assume that no other cause is operative – that the effect is not overdetermined. But this assumption should not be built into the definition of causation. For it is an additional assumption that people commonly make when they assert that E₁ causes E₂, and not a necessary condition for the truth of that assertion.

The common assumption that no other cause is operative may lie behind the puzzling remark of Hume with which we began this part. Hume, you will recall, seems mistakenly to equate "*all the objects similar to the first are followed by objects similar to the second*" with "*if the first object had not been, the second had never existed.*" But perhaps his thought is this: when we assert that E₁ caused E₂, this means that (a) E₁ was sufficient for E₂. However, we commonly make two additional assumptions: (b) that E₂ would not have occurred without *some* sufficient condition (other than itself); and (c) that there was no *other* sufficient condition for E₂ besides E₁. Now from (a), (b), and (c), it does follow that *if E₁ had not occurred, E₂ would not have occurred*; or, in Hume's words, that "*if the first object had not been, the second never had existed.*" Thus, we can understand why Hume might have thought that this remark was an appropriate way for him to paraphrase or gloss his definition of causation.

In order to avoid the difficulties we have explored in this part, however, we need to see that (c) does not have to be true in order for "E₁ caused E₂" to be true. Therefore, (c) should not be included in the definition or analysis of causation. It should be regarded, as suggested above, as an additional assumption that we generally make when we assert that E₁ caused E₂. As for (b), it would seem that any analysis of causation which says that a cause, or some member of a disjunction of possible causes, is a necessary condition for its effect, must hold that (b) is part of the meaning of "E₁ causes E₂." Thus, such analyses seem to imply that any event that was caused could not have occurred uncaused. Provided that this "could not" is not interpreted in the sense of logical impossibility, this implication may well be acceptable. But if the implication is unacceptable – a matter that we shall not try to settle – then it is open to the regularity theorist to jettison the idea that a cause (or some member of a disjunction of possible causes) is a necessary condition for its effect, by sticking with the unexpanded version of *D3* that uses 2 or 2A rather than 2" or 2A".

The problem of collateral effects

The final problem for the regularity theory that we shall consider is a simple yet perplexing one, which has been emphasized by J. L. Mackie and other contemporary philosophers. Consider a cause, C, with two collateral effects, E₁ and E₂, where E₁ precedes E₂. How can a regularity theorist avoid saying, falsely, that E₁ causes E₂? To make this problem clearer,

we can use an example given by Bernard Berofsky (1983: 485), in a review of a recent book that offers a detailed defense of the regularity theory. Suppose that "... a fatal viral disease V produces a distinctive rash R . . ." In such a case, the virus causes both the rash and the death, but the rash does not cause the death. As Berofsky puts it, "the sequence $V \rightarrow$ Death is causal, whereas the sequence $R \rightarrow$ Death is not." The problem is: how can a regularity theorist avoid saying, falsely, that R causes the death?

A regularity theorist could reply as follows. The problem turns on the assumption that R is always followed by death. So, according to the regularity theory, it is a law of nature that *whenever* R occurs, death occurs; in which case R causes death. However, the example does not really show that whenever R occurs, death occurs. Rather, it shows only that whenever R occurs *and was preceded by* V , death occurs. But, as we emphasized in the previous part, a given effect can be caused in a variety of different ways. Thus, R could be induced by an experimental vaccine or a drug, or by exposure to some irritant. But then R would not be followed by death; so "whenever R occurs, death occurs" is not a law of nature after all.

A critic of the regularity theory could make two possible responses to this reply. One response would be to stipulate that if the rash is artificially induced or caused in any way other than by V , then it is not a rash of kind R . Now this would be to say that being accompanied by V is part of what it means to be a rash of kind R , so that from the fact that a rash is of kind R , it follows logically that it is accompanied by V . But in that case, there are two points that a regularity theorist can make. First, we cannot say that R is caused by V . For, as Hume showed, one cannot logically deduce that a given cause has occurred solely from knowledge of its effect, any more than one can logically deduce that a given effect will occur solely from knowledge of its cause. Second, if part of what it means to be a rash of kind R is to be accompanied by V , then there is no reason to deny that R is a cause of death. For we know that V is a cause of death, and V is now, by stipulation, a constituent element of R .

The critic's other possible response would be to reject the assumption that there are other ways to produce R . Although all the rashes that we know about can be caused in a variety of different ways, the critic could stipulate that R is a very special rash that occurs only in connection with V : it cannot be artificially induced or produced in any other way. Of course, V would then be both sufficient and necessary for R ; so it might seem that the asymmetry needed for V to be the cause of R would be absent. However, the critic could also stipulate that V precedes R in time, thereby preserving the asymmetry.

The regularity theorist, however, could now reply that it is no longer clear that R is not a cause of death. If such a case occurred, we might say that both R and V were causes of death, or perhaps that the complex event V -and- R was a cause of death. Of course, it sounds odd to say that the rash is just as much a cause of death as is the virus. But this may be because in all cases of fatal viruses that we know about, only the virus itself is the cause of death. Symptoms such as rashes and the like, since they can also occur without the fatal virus, are not causes of death. Nor do they occupy the central place in a theory of disease that the viruses occupy. But if the situation described by the critic were actual – if R were necessary

and sufficient for V and R -and- V were sufficient for death – then R would occupy just as central a place in the diagnosis and treatment of the particular disease involved as does V , and it seems that there would be no reason to exclude its consideration as a cause of death.

HUME'S CRITIQUE OF THE CAUSAL PRINCIPLE

1 Introduction

Virtually all of us accept the principle that every event has a cause, even if we do not always know what that cause is. The idea of an uncaused event – of something happening without any cause that explains why it happened – strikes us as somehow “irrational” or absurd. As for philosophers, many of them have not only accepted the truth of the principle that every event has a cause, but have assumed that this truth is knowable by any rational mind, in much the same way as a simple mathematical truth like $2 + 2 = 4$ or a simple logical truth like *not both p and not-p*. In other words, many philosophers have assumed that the causal principle (as we shall call the principle that every event has a cause) is knowable a priori. Accordingly, they have often gone on to use this principle in order to “prove” important conclusions, notably the existence of God. The principle has been used in this way by, for example, Aquinas, Descartes, Locke, and Berkeley, to name just a few.

Contrary to what is sometimes thought, Hume also accepts the causal principle as a true proposition. In a discussion of “freedom” in the *Enquiry*, he writes: “it is universally allowed that nothing exists without a cause of its existence” (E:95; S:63–4; F:131). He does not demur from this principle, but instead goes on to argue that it is compatible with human freedom. And in one of his letters Hume says: “I never asserted so absurd a Proposition as *that anything might arise without a Cause . . .*” (Greig 1932: 187). What Hume does not accept, however, is the view that the causal principle is knowable a priori; or, in his terminology, that it is either “intuitively or demonstratively certain” (self-evident or demonstrable).¹ In other words, Hume agrees with most other philosophers that the causal principle is *true*, but he disagrees with philosophers who came before him about how it can be *known to be true*; for *they* held it to be a priori, whereas Hume argues, famously, that it is neither self-evident nor demonstrable. Instead, Hume maintains that the causal principle is merely an empirical generalization: “the opinion of the necessity of a cause to every new production . . . must necessarily arise from observation and experience” (T:82; F:50).

Already, in previous chapters, we have reconstructed and favorably evaluated Humean arguments showing that no bridging proposition is self-evident or demonstrable, and pointed out that those arguments apply to the causal principle.² Hume, however, offers also arguments designed specifically to show that the causal principle is neither self-evident nor demonstrable, and it is to those arguments that we turn in this chapter. Hume's arguments themselves are analyzed in part 2. In part 3 we shall examine one contemporary critique of Hume's arguments, and lastly we shall consider Immanuel Kant's famous attempt to "answer" Hume's arguments.

2 Why the causal principle is neither self-evident nor demonstrable

Hume's critique of the causal principle is given only in the *Treatise*, Book I, Part iii, Section 3, which is entitled, ironically, "Why a Cause Is Always Necessary." Hume begins this section (hereafter called simply "Section 3") by accurately remarking that the causal principle was

commonly taken for granted in all reasonings, without any proof given or demanded. [For it was] suppos'd to be founded on intuition, and to be one of those maxims, which tho' they may be deny'd with the lips, 'tis impossible for men in their hearts really to doubt of.

(T:79; F:47)

In the discussion that follows, Hume formulates the principle in a number of different ways. Two of these are that "*whatever begins to exist, must have a cause of existence*" (T:78) and that "*whatever has a beginning has also a cause of existence*" (T:79; F:47). Those formulations make it sound as if Hume were concerned only with the causes of things' existence; however, Hume speaks also of our inability to "demonstrate the necessity of a cause to every new existence, *or new modification of existence*" (T:79; F:47; emphasis added), thereby showing that his concern extends to the causes of any event whatsoever (see Penelhum 1992: 117). His arguments, however, are couched in terms of the causes of things' beginning to exist, and we shall reconstruct them in those terms in order to stay close to his text.

Hume argues first that the causal principle is not self-evident ("intuitively certain"). His argument turns on a theory about relations that appears in the *Treatise* (see Book I, part iii, Section 1) but is dropped in the *Enquiry*, where it is replaced by the fundamental distinction between relations of ideas and matters of fact. The *Treatise's* theory implies that a self-evident proposition must owe its self-evidence to one of four possible relationships between its terms, which Hume calls "*resemblance, proportions in quantity and number, degrees of any quality, and contrariety*" (T:79; F:47). Thus, for example, the statement that "red is more like pink than like white" can be known just by understanding what it says, because anyone

who understands it grasps that the resemblance relations it reports are just what it states them to be; and the statement “no object is both black and white all over” is self-evident because anyone who understands it grasps that “black” and “white” are contrary terms. Hume then concludes that the causal principle is not self-evident, since it does not turn on any one of the four relations just mentioned.

A full-scale evaluation of this argument would require asking whether the four relations Hume specifies are the only possible sources of self-evidence. That there are only these four seems doubtful: the statement “if I am thinking, then I exist,” for example, is self-evident, but seems not to involve any of Hume’s four relations. Perhaps Hume himself came to think that his argument was a weak one, since he drops the *Treatise*’s theory of relations in the *Enquiry*. Be that as it may, the argument’s weakness seems hardly to matter. For on reflection, it seems sufficiently clear that the causal principle, regardless of which of Hume’s formulations we focus on, is not a proposition that can be known to be true merely by understanding what it says.

Next, Hume offers, in Section 3’s third paragraph, a highly condensed and complex argument, which, he says, “proves at once that [the causal principle] is neither intuitively nor demonstratively certain.” Here is the entire paragraph:

But here is an argument, which proves at once, that the foregoing proposition is neither intuitively nor demonstrably certain. We can never demonstrate the necessity of a cause to every new existence, or new modification of existence, without shewing at the same time the impossibility there is, that any thing can ever begin to exist without some productive principle; and where the latter proposition cannot be prov’d, we must despair of ever being able to prove the former. Now that the latter proposition is utterly incapable of a demonstrative proof, we may satisfy ourselves by considering, that as all distinct ideas are separable from each other, and as the ideas of cause and effect are evidently distinct, ’twill be easy for us to conceive any object to be non-existent this moment, and existent the next, without conjoining to it the distinct idea of a cause or productive principle. The separation, therefore, of the idea of a cause from that of a beginning of existence, is plainly possible for the imagination; and consequently the actual separation of these objects is so far possible, that it implies no contradiction or absurdity; and is therefore incapable of being refuted by any reasoning from mere ideas; without which ’tis impossible to demonstrate the necessity of a cause.

(T:79–80; F:47–8)

Let us grant that the causal principle is not self-evident (i.e. not “intuitively certain”), and try to reconstruct the argument showing that it is not demonstrable (i.e. not “demonstratively certain”), by analyzing the above paragraph into slightly edited segments and considering each in turn.

The first segment goes this way:

We can never demonstrate the necessity of a cause to every new existence, or new modification of existence, without shewing at the same time the impossibility there is, that any thing can ever begin to exist without some productive principle; and where the latter proposition cannot be prov'd, we must despair of ever being able to prove the former. Now . . . the latter proposition is utterly incapable of a demonstrative proof.

Hume's reasoning here is very straightforward: we cannot prove that every beginning of existence must have a cause unless we can prove that a beginning of existence without a cause is impossible; but we cannot prove that a beginning of existence without a cause is impossible; *ergo*, we cannot prove that every beginning of existence must have a cause.

But why does Hume think that we cannot prove the impossibility of a beginning of existence without a cause? The answer begins in the next segment, where Hume undertakes to show why "the latter proposition [i.e. that a beginning of existence without a cause is impossible] is utterly incapable of a demonstrative proof" and which is, from a logical point of view, the real starting-point of the argument (I have inserted numbers to identify the steps of the reasoning):

[A]s (1) all distinct ideas are separable from each other, and as (2) the ideas of cause and effect are evidently distinct, (3) 'twill be easy for us to conceive any object to be non-existent this moment, and existent the next, without conjoining to it the distinct idea of a cause or productive principle.

Step (2) is carelessly stated; for Hume is trying to show in step (3) that we can conceive a beginning of existence (an object that is "non-existent this moment, and existent the next") without a cause of existence – that we can conceive a certain kind of *event* occurring without any cause; not that we can conceive an *effect* without a cause, which would be absurd. Correcting for this carelessness, then, we can formulate the first part of Hume's argument like this:

- (1) All distinct ideas are separable from each other (premiss).
- (2) The idea of a cause of existence is distinct from the idea of a beginning of existence (premiss).
- (3) We can conceive of something beginning to exist without a cause (from (1) & (2)).

Hume states the third part of the argument in the final segment:

The separation, therefore, of the idea of a cause from that of a beginning of existence, is plainly possible for the imagination; and consequently the actual separation of these objects is so far possible, that it implies no contradiction or absurdity; and is

therefore incapable of being refuted by any reasoning from mere ideas; without which 'tis impossible to demonstrate the necessity of a cause.

Here it would seem that the first clause (up to the semicolon) is merely a restatement of (3) – the claim that we can *conceive* a beginning of existence without a cause – despite the fact that Hume switches from talk about what we can conceive to talk about what is “possible for the imagination.” He moves now from this claim to the further claim that there is no “contradiction or absurdity” in the notion of a causeless beginning of existence, and from that claim to his key claim that we cannot “refute” the possibility of such a beginning of existence. To make these inferences, however, two premisses that Hume leaves unstated are required: namely, a premiss to the effect that whatever we can conceive or “imagine” is free of contradiction, and a premiss to the effect that whatever is free of contradiction cannot be proved to be impossible. Thus, we can formulate this part of the argument as follows:

- (4) Nothing that we can conceive implies a contradiction (suppressed premiss).
- (5) “*X began to exist and X had no cause*” does not imply a contradiction (from (3) & (4)).
- (6) If *p* does not imply a contradiction, then we cannot demonstrate that *p* is impossible (suppressed premiss).
- (7) We cannot demonstrate that a beginning of existence without a cause of existence is impossible (from (5) & (6)).

Hume indicates that his argument is now substantially complete by closing the paragraph with the words, “without which 'tis impossible to demonstrate the necessity of a cause.” This is just a reminder of the reasoning in the paragraph's first segment, which we can now enter into our reconstruction of the argument in its logically correct place, thereby also completing the reconstruction:

- (8) We can demonstrate that whatever has a beginning of existence must have a cause of existence only if we can demonstrate that a beginning of existence without a cause of existence is impossible (premiss).
- (9) We cannot demonstrate that whatever has a beginning of existence must have a cause of existence (from (7) & (8)).

Having given this complex argument, Hume goes on to confirm its result by refuting three attempts to prove the causal principle. The three attempted proofs and Hume's refutations of them are much simpler than is the argument just analyzed, but highly instructive nevertheless. Let us consider them.

The first attempt, which Hume attributes to Thomas Hobbes, is this:

All the points of time and place . . . in which we can suppose any object to begin to exist, are in themselves equal; and unless there be some cause, which is peculiar to one time and to one place, and which by that means determines and fixes its existence, it must remain in eternal suspense; and the object can never begin to be, for want of something to fix its beginning.

(T:80; F:48)

This argument is a syllogism that can be put this way:

- (1) All beginnings of existence must be beginnings of existence at a particular place and time.
 - (2) All beginnings of existence at a particular place and time must have a cause.
-

∴ All beginnings of existence must have a cause.

The nerve of the argument is (2), which invokes the idea that there must be a cause for a thing beginning to exist at one particular place and time rather than another. But Hume responds by rejecting (2). He argues that if the denial of (3) is not “intuitively absurd” – that is, if (3) is not self-evident – then the denial of (2) is not “intuitively absurd” either – that is, (2) is not self-evident either. Thus, for example, if it is not intuitively absurd for a unicorn to begin to exist without any cause, then neither is it intuitively absurd for a unicorn to begin to exist at noon on January 1, 2000 in San Diego Zoo, rather than at midnight on January 2, 2000 in London Zoo, without any cause. Hume’s point is basically that if (3) is regarded as needing a proof, then (2) must be regarded also as needing a proof; and so (2) cannot be used as an unsupported premiss in an argument for (3).

The second attempt, which Hume attributes to Samuel Clarke, goes this way: “Every thing . . . must have a cause; for if anything wanted a cause, *it* would produce *itself*; that is, exist before it existed; which is impossible” (T:80; F:48). This argument can be stated as follows:

- (1) If a thing has no cause, then it must produce itself.
 - (2) If a thing produces itself, then it must exist before it exists.
 - (3) Nothing can exist before it exists.
-

∴ Everything must have a cause.

Notice that (1), as stated, is very confusing; for it seems to be saying that if a thing has no cause, then it does have a cause after all! (Namely, itself.) Hume’s trenchant response is that the argument simply begs the question; for (1) can only be a confused way of saying that if a thing does not have some *other* thing for a cause, then it must have *itself* for cause – which is to assume the very point in question; namely, that everything must have a cause. As Hume puts it:

[T]o say that anything is produc'd, or to express myself more properly, comes into being, without a cause, is not to affirm, that 'tis itself its own cause; but on the contrary in excluding all external causes, excludes *a fortiori* the thing itself, which is created. An object, that exists absolutely without any cause, certainly is not its own cause; and when you assert, that the one follows from the other, you suppose the very point in question, and take it for granted, that 'tis utterly impossible any thing can ever begin to exist without a cause, but that upon the exclusion of one productive principle, we must still have recourse to another.

(T:81; F:49)

The third attempt, which Hume attributes to John Locke,³ is this:

Whatever is produc'd without any cause, is produc'd by *nothing*; or in other words, has nothing for its cause. But nothing can never be a cause, no more than it can be something, or equal to two right angles. By the same intuition, that we perceive nothing not to be equal to two right angles, or not to be something, we perceive, that it can never be a cause; and consequently must perceive, that every object has a real cause of its existence.

(T:81; F:49)

The argument can be put like this:

- (1) If X is produced without any cause, then X is produced by nothing (i.e. X has nothing for its cause).
- (2) Nothing cannot be a cause (any more than it can be something, or can be equal to two right angles).

∴ Everything must have a cause.

Notice that, like the first premiss of Samuel Clarke's argument, (1) is stated in a confused manner; for it seems, again, to be saying that if X is produced without any cause, then it has a cause after all – namely, nothingness, or non-being. Hume responds that the argument begs the question in almost exactly the same way as does Clarke's argument; for its first premiss can only be a confused way of saying that if X does not have some *existing* thing for a cause, then it must have nothing (nothingness, non-being) for a cause – which again is to assume that X must have a cause.

Hume's criticisms of Clarke and Locke's arguments are very instructive. For they show that the statement *Maybe X has no cause* does not mean, as Locke thinks, (a) Maybe X is caused by nothing(ness) or non-being; or, as Clarke thinks, (b) Maybe X is its own cause.

Rather, it means only (c) Maybe X is *without* any cause. So the absurdity of (a) and (b), which Hume does not dispute, does not show that (c) is absurd.

Hume's criticism of Locke's argument could be applied also to this simple attempt to demonstrate the causal principle, which might at first seem quite plausible:

(1) Something cannot come from nothing (*Ex nihilo, nihil fit*).

∴ Everything must have a cause.

A person who has understood Hume's response to Locke's argument could point out that the premiss, (1), has two possible meanings:

- (1a) Nothing(ness) or non-being cannot be a cause; that is, something cannot be caused by nothing(ness) or non-being.
- (1b) Something cannot exist without any cause; in other words, something cannot lack a cause.

If we suppose that the premiss "Something cannot come from nothing" means (1a), then (1) is self-evident; but (1) does not entail (2). On the other hand, if we suppose that (1) means (1b), then (1) *does* entail (2); indeed (1) and (2) are then logically equivalent. But if (1) means (1b), then (1) is not self-evident. The result is that on either interpretation of its premiss, this argument fails to prove the causal principle.

3 Stroud's critique of Hume's argument

In his book *Hume*, Barry Stroud criticizes Hume's complex proof that the causal principle cannot be demonstrated. In this part, we shall argue that Stroud's objections do not refute Hume's argument.

It will be helpful to begin by reviewing the steps of Hume's argument:

- (1) All distinct ideas are separable from each other (premiss).
- (2) The idea of a cause of existence is a distinct idea from the idea of a beginning of existence (premiss).
- (3) We can conceive of something beginning to exist without a cause (from (1) & (2)).
- (4) Nothing that we can conceive implies a contradiction (suppressed premiss).
- (5) "X began to exist and X had no cause" does not imply a contradiction (from (3) & (4)).
- (6) If *p* does not imply a contradiction, then we cannot demonstrate that *p* is impossible (suppressed premiss).
- (7) We cannot demonstrate that a beginning of existence without a cause of existence is impossible (from (5) & (6)).

- (8) We can demonstrate that whatever has a beginning of existence must have a cause of existence only if we can demonstrate the impossibility of a beginning of existence without a cause of existence (premiss).
- (9) We cannot demonstrate that whatever has a beginning of existence must have a cause of existence (from (7) & (8)).

Stroud (1977: 46–50) makes two objections to this argument: first, that Hume's justification for (7) is circular; second, that (4) is simply false. Let us consider the objections in turn.

To show that Hume's justification for (7) is circular, Stroud begins by asking: how does Hume know that (2) is true – that the idea of a cause of existence is an idea different or distinct from the idea of a beginning of existence? It might be answered, says Stroud, that we know this by seeing that a beginning of existence without a cause of existence doesn't imply a contradiction. But in that case, he objects, (5) is being used to support (2), and so (2) cannot be used to support (5) without arguing in a circle. Furthermore, now we need some other way to support (5). It will not do, of course, to say that (5) is known to be true because, for all we can know a priori, an uncaused beginning of existence *is* possible. For this would be to use (7) to support (5), in which case, Stroud says (p. 48), "Hume's 'argument' . . . would shrink to the mere assertion of the possibility of something's beginning to exist without a cause."

Stroud next considers a way in which it may be possible to show that (5) is true without appealing to (7). To show this, it is necessary to provide some way of deciding whether a statement is a contradiction. If we confine our attention to statements formulated in the constructed, artificial language of formal logic, it is not difficult to provide such a test: a statement is a contradiction if and only if it has the form "p and not-p." This is something that we can determine just by looking at the statement. But when we broaden our attention to statements in a natural language like English, we are not always able to tell whether a statement is a contradiction just by looking at it. The statement, "there is a husband who lacks a wife," for example, is not obviously of the form "p and not-p;" it seems rather to have the form "p and not-q."⁴ Some philosophers would give the following test for deciding when a statement in a natural language is a contradiction: a statement S is a contradiction just in case, when we replace some term(s) in S with other terms that stand for the same idea, the resulting statement has the form "p and not-p." For example, "there is a husband who lacks a wife" *is* a contradiction, because once we replace "husband" with "man who has a wife," then we see that the statement says, "there is a man who has a wife and does not have a wife" – which has the form "p and not-p."⁵ Now, it might be said that applying this test to "X began to exist and X had no cause" shows that this statement does not imply a contradiction. For we cannot substitute terms in this statement with other terms that have the same meaning and thereby obtain a statement of the form "p and not-p." And this is why "X began to exist and X had no cause" does not imply a contradiction – which is to say, why (5) is true.

As we might expect in light of Stroud's initial objection, however, he rejects the proposed test for deciding when a natural-language statement implies a contradiction. For the proposed

test obviously “makes essential use of the notion of the ‘same’ or ‘distinct’ ideas” (Stroud 1977: 48). So it is circular to use the proposed test as the test for contradictoriness, since contradictoriness is itself the only test that has so far been offered for the notion of “same/distinct idea.” In other words, (2) is again being used to support (5), despite the fact that the only support given for (2) has been (5). Stroud concludes that “Hume really has no non-circular argument on this point at all. He thinks he can start from the ‘evident’ distinctness of two ideas, but he never says how he can recognize that distinctness” (ibid.).

Despite Stroud’s ingenuity, his contention is debatable. He has certainly shown that sameness/distinctness of ideas, conceivability, contradiction, and possibility are interdependent notions; much as W. V. O. Quine, whom Stroud cites as an inspiration for his objection, does with analyticity, sameness of meaning, and synonymy in his famous paper “Two Dogmas of Empiricism” (Quine 1953: 20–46). It is questionable, however, that there is a vicious circularity in using some of these notions to clarify the others. After all, the notions of “same” and “distinct” ideas are not totally obscure; and in the case at hand it might even be plausibly said that the idea of a *beginning* of existence is clearly distinct from that of a *cause* of existence (as indeed Hume seems to believe); so that one can argue from this distinctness to the conceivability of an uncaused beginning of existence and from thence to the latter’s possibility, as Hume does when he goes from (2) to (3) to (5). But if someone wanted confirmation of the distinctness of these two ideas, we could reverse the argument, by pointing out that it is conceivable that a thing might spring into existence *ex nihilo*, or that this supposition implies no contradiction. Hume’s case for (7) would be viciously circular only if none of (2) or (3) or (5) had any independent plausibility; but in fact each of them is independently plausible, and so steps (1)–(7) are better seen as exhibiting the interconnections between the notions involved than as a linear defense of (7).

Stroud’s second objection to Hume’s argument is that (4) – the premiss that nothing of which we can conceive implies a contradiction – is false. In other words, conceivability is not a decisive criterion of what does or does not imply a contradiction (or of what is or is not impossible). To show this, Stroud points out that Goldbach’s conjecture – that every even number is the sum of two prime numbers – has never been proved or disproved. Yet the conjecture is either necessarily true (in which case its denial implies a contradiction and is impossible), or necessarily false (in which case its affirmation implies a contradiction and is impossible). Now, says Stroud (p. 50),

it seems easy to conceive of Goldbach’s Conjecture’s being proved one day . . . But I can also conceive of its being disproved, of someone’s proving its negation, perhaps by finding a very large even number that is not the sum of two primes. I can conceive of a computer’s coming up with one tomorrow.

Therefore, I can conceive of something that implies a contradiction and is impossible. So, (4) is false.

This objection seems flawed, also. For Stroud’s conceivability claims seem to mean only

that one can imagine someone doing some calculations and announcing that they prove Goldbach's conjecture, or imagine a person or a computer identifying a very large even number and telling us that it is not the sum of two primes. But it is questionable that this amounts to conceiving of the truth or the falsity of Goldbach's conjecture itself. After all, one can imagine or "conceive" of a famous mathematician declaring that he has proved that $1 + 1 = 3$, or of a computer outputting that $1 + 1 = 3$; but it does not follow that one can conceive that $1 + 1 = 3$.

Finally, we should point out that even if Stroud's objections are thought to carry some weight against Hume's explicit argument, it is possible to give a simpler proof that the causal principle is not demonstrable, based on "principle (P)" (see Chapter 2 part 5, and Chapter 3 part 3). This simpler argument is the following:

- (1) If p is demonstrable, then there is a set of statements, S , such that (i) all the members of S are self-evident, and (ii) not- p , together with S , entails a contradiction (principle (P)).
 - (2) There is no set of statements, S , such that (i) all the members of S are self-evident, and (ii) "X began to exist and X had no cause," together with S , entails a contradiction.
-
- ∴ (3) "Whatever has a beginning of existence has a cause of existence" is not demonstrable.

Given Stroud's objections to the use of the notion of contradiction (outside the context of purely formal logic), it would be disingenuous to claim that he must accept this argument even if he rejects Hume's own explicit argument. But if we are willing to accept the notion of contradiction, then this essentially Humean argument seems very powerful.

4 Kant's "answer" to Hume

The most famous attempt to answer Hume's arguments against the demonstrability of the causal principle is that of Immanuel Kant. This "answer" is given in a section of his *Critique of Pure Reason* called the "Second Analogy of Experience" (hereafter referred to as "the second analogy"). Kant's discussion of the causal principle is embedded in a very difficult and complex theory of knowledge, and there is a vast literature on that theory as well as specifically on Kant's second analogy. Although we cannot possibly do justice to Kant's discussion, its importance and direct bearing on Hume's treatment of the causal principle warrant giving an account of it here.

Although Kant disagrees with Hume about the status of the causal principle, it is important to note at the outset some points of agreement between the two. First, Kant agrees with Hume that the causal principle is not true simply by virtue of conceptual relationships or meanings of words – that it is not, in the modern parlance developed by Kant himself, an analytic proposition, but rather a synthetic one. Second, Kant does not think any more than

does Hume that the causal principle can be demonstrated by manipulating general concepts like existence, beginning of existence, event, cause, and so on. Rather, Kant tries to show that the principle can be proved through what he calls a “transcendental” argument. This is an argument according to which the truth of a certain principle (in this case the causal principle) is a necessary condition of experience. So Kant’s argument for the causal principle is that unless that principle were true, we could not have the sort of experience that we do. In order to appreciate how such an argument works, it is crucial to know exactly how the term “experience” is to be understood. In his *Critique of Pure Reason*, Kant gives a number of transcendental arguments for a number of different principles, and the term “experience” seems not to mean exactly the same thing in all of those arguments. But in the argument of the second analogy, the kind of experience Kant has in mind is that of knowing by perception that an event has occurred. Kant, as we shall see, tries to show that unless the causal principle were true, we could never know by perception that any event had occurred.

According to Kant, the fact that the causal principle can be proved only by a transcendental argument has a very important consequence: namely, that the principle can be proved to hold only for *observable* events; it cannot be proved to hold for events that we could not possibly experience. This means that there is an additional similarity between Hume and Kant’s views of the causal principle: namely, that it cannot be used to show that events which supposedly occur totally outside the field of our experience, such as the origin of the universe, must have a cause. Thus, even if the principle can be demonstrated, it cannot be used, in the way that Rationalist philosophers employed it, to establish the existence of God or other entities that could not fall within the scope of our experience.

Having mentioned some similarities between Hume and Kant’s views of the causal principle, we shall henceforth focus on the differences. First, there is an important difference between the version of the causal principle that Hume says cannot be proved and the version Kant thinks he can establish. The version Hume considers is unrestricted: it would apply to any beginning of existence or “modification of existence” whatsoever; simply put, Hume’s version says that *every event whatsoever* has a cause. Now the principle that Kant claims to prove in his second analogy is this: “Everything that happens, that is, begins to be, presupposes something on which it follows according to a rule” (Kant 1963: A189). The wording of this principle makes it sound very much like the principle that Hume says cannot be proved: “whatever has a beginning of existence has a cause of existence.” Despite its wording, however, Kant’s principle contains the restriction just mentioned: it applies only to *observable events*. In other words, Kant thinks that he can prove that every event that we could ever observe must have a cause, but not that every event, period, must have a cause. However, Hume’s arguments imply that not even Kant’s restricted version of the causal principle can be proved; so a chief difference between Hume and Kant is that Hume holds that no version of the causal principle can be demonstrated, whereas Kant argues that a version restricted to observable events can be demonstrated. By showing that the version of the causal principle which is restricted to observable events can be demonstrated, while maintaining that an unrestricted version, supposedly applying beyond the bounds of human experience, cannot

be known to be true, Kant saw himself as defending the foundations of Newtonian physics, without lapsing into the Rationalist metaphysics that he rejected no less than Hume did. Finally, Kant's "transcendental" way of arguing for the causal principle implies another difference between his position and Hume's, which is well stated by William H. Brenner:

Hume thought that the principle of causality was a generalization from our experience of events. But if Kant's argument is sound, then all perception of events, and consequently all generalization from experience, *presupposes* the principle of causality. Kant's answer to Hume . . . is that the principle of causality is *presupposed* by the perception of events, not *derived from* it.

(1989: 128)

In order for Kant to give an argument in support of (even his restricted version of) the causal principle that can "answer" Hume, that argument must start from premisses Hume himself would accept. Now, as the Kant scholar Lewis White Beck has shown (1978: 130–5) in an effective analysis of Kant's strategy for answering Hume, Kant's argument in the second analogy does start from a point which is common ground for them both. This is that any knowledge we have of causal relationships must be based on induction: we know that A-events cause B-events only because, in all cases we have observed, A-events have been followed by B-events – because A-events and B-events have been constantly conjoined in our experience.⁶ As Beck says:

[Kant] is in complete agreement with Hume that our knowledge of causal connections between specific events is a posteriori not a priori, synthetic not analytic, inductive not logical, probable not certain. His methods for finding the cause of B are exactly those which Hume prescribed, and the chances of success in this venture, as estimated by Kant and Hume, are very much the same. Kant's first answer to Hume, then, is to agree with him, and to disagree with the rationalists who thought that logical insight into causal connections was possible.

(Beck 1978: 134)

As Beck points out, however, the ability to infer that A-events cause B-events from observing that A-events have been regularly followed by B-events presupposes that we can identify or discriminate events perceptually, that is, that we can tell by observation that an event is occurring. For if we could not do this, then we could not establish the *premiss* of the inference – that *events* of a certain kind have been regularly followed by *events* of a certain other kind.

Now Kant's key insight in the second analogy is that there is an epistemological problem about how we are able perceptually to identify or discriminate events. Specifically, there is

a problem about how we are able to distinguish events from enduring states of affairs; for, whether we are perceiving an event or an enduring state of affairs, our perceptions occur successively or serially in time. Kant illustrates this point with the examples of perceiving a ship moving downstream and of perceiving a house. The ship moving from an upstream position to one further downstream is an event. On the other hand, the existence of the various parts of the house – its front, sides, back, foundation, roof – is an enduring state of affairs. But, in both cases, our perceptions occur successively or serially in time. In the case of the ship, we see it first upstream and then downstream. In the case of the house, we see first one side and then another side, or first the foundation and then the roof, or first the roof and then the foundation. This shows that we cannot tell, merely from the fact that our perceptions occur serially or successively, that we are perceiving an event rather than an enduring state of affairs. In other words, observation alone provides no criterion by which we can distinguish the two. How, then, can we tell if we are perceiving an event or an enduring state of affairs? To quote Beck: “[Hume] never discussed this problem; no one before Kant even saw that it was a problem” (1978: 135). Kant’s thesis in the second analogy is that this problem can be solved in only one way: namely, if we grant that every observable event has a cause, or, as Kant puts it, that “everything that happens, that is, begins to be, presupposes something on which it follows according to a rule.” In other words, Kant contends that we can distinguish between events and enduring states of affairs, and so identify events, only if the causal principle is true of those events.

Beck gives a succinct summary of Kant’s strategy, involving three propositions (the wording of which has been slightly modified):

- H* From observing repeated pairs of similar events, we infer that events like the first members of the pairs are causes of events like the second.
- P* Events can be distinguished from enduring states of affairs, even though our perceptions of both are successive or serial.
- K* “Everything that happens, that is, begins to be, presupposes something on which it follows by rule” (ibid.: 135).

Proposition *H* is common ground between Hume and Kant: thus it is a premiss that Hume himself accepts and is in no way question-begging.⁷ Proposition *K* is Kant’s statement of the causal principle which Hume says cannot be demonstrated. Clearly *H* implies *P*: we cannot establish correlations between events unless we can distinguish events from enduring states of affairs. The task of the second analogy is to show that *P* in turn implies *K*. If that can be shown, it will follow that *H* implies *K*, and thus that the causal principle can be demonstrated from a premiss which Hume himself accepts (Beck 1978: 135).

In what follows, we shall consider two possible arguments for getting from *P* to *K*. The first is quite strongly suggested by the text of the second analogy, and has often been thought to be Kant’s authentic argument. However, as P. F. Strawson shows in his influential book on

Kant *The Bounds of Sense* (1966: 133–40), it is a fallacious argument. The second argument is also suggested by Kant's text, and seems more promising.

The first argument, which I call the "irreversibility" argument, is suggested by what Kant says when he compares the examples of the house and the ship. In the case of the house, the series of perceptions obtained by the observer may be said to be *reversible*. This is because, depending on the circumstances and on the way the observer chooses to view the house, the observer can see first the front of the house and then the back of the house, or first the back and then the front; likewise the observer can see first the left side and then the right side or first the right side and then the left side, and first the basement and then the roof or first the roof and then the basement. In other words, in whatever order the observer's perceptions occur, they could have occurred in the opposite or reverse order instead. Kant puts the point this way:

In the . . . example of a house my perceptions could begin with the apprehension of the roof and end with the basement, or could begin from below and end above; and I could similarly apprehend the manifold of empirical intuition either from right to left or left to right. In the series of perceptions there was thus no determinate order specifying at what point I must begin in order to connect the manifold empirically.⁸

(1963: A192–3)

In the case of the ship moving downstream, on the other hand, the series of perceptions may be said to be *irreversible*. Assuming that the ship is moving downstream, one's perceptions can occur in only one order: first one sees the ship upstream and then one sees it downstream; one cannot see it first downstream and then upstream. One's perceptions cannot occur in any order other than the one that corresponds to the ship's successive positions in the stream. As Kant puts it:

But, as I also note, in an appearance which contains a happening (the preceding state of the perception we may entitle A, and the succeeding B) B can be apprehended only as following upon A; the perception of A cannot follow upon B but only precede it. For instance, I see a ship move down stream. My perception of its lower position follows upon the perception of its position higher up in the stream, and it is impossible that in the apprehension of this appearance the ship should first be perceived lower down in the stream and afterwards higher up. The order in which the perceptions succeed one another in apprehension is in this instance determined, and to this order apprehension is bound down.

(1963: A192)

At the end of the same paragraph, Kant concludes that “in the perception of an event there is always a rule that makes the order in which the perceptions (in the apprehension of this appearance) follow upon one another a *necessary* order” (ibid.: A193).

It is chiefly from these passages that the irreversibility argument is drawn. Strawson in effect divides the argument into two stages. In the first stage, Kant is seen as pointing to a criterion by which our perceptions of an event can be distinguished from those of an enduring state of affairs, despite the perceptions of both being successive or serial. This criterion is the reversibility or irreversibility of the series of perceptions. Thus the criterion by which a series of perceptions is apprehended as of (or taken to be of) an enduring state of affairs is *reversibility*: the series could have been obtained in the reverse order from that in which it actually occurred, as in the case of the house. And the criterion by which a series of perceptions is apprehended as of (or taken to be of) an event is *irreversibility*: the series could not have been obtained in the reverse order from that in which it actually occurred, as in the case of the ship. In other words, perceptions are enduring state of affairs if and only if they are reversible; whereas perceptions are of an event if and only if they are irreversible.⁹ In the second stage of the argument, Kant is seen to argue from the irreversibility of perceptions of an event to the truth of the causal principle: since our perceptions of events are irreversible, those events must be subsumed under causal laws.

In order to evaluate this well-known yet puzzling argument, we need to state it in a somewhat more formal way. From the first stage of the argument, in which the reversibility–irreversibility criterion is put forward, we can extract the following premiss:

- (1) Necessarily, if S perceives an event $A-B$, then S's perceptions occur in the order A, B .

Here “an event $A-B$ ” means an event or change whose first stage is A and whose second stage is B ; so that in Kant's example of the moving ship, A would designate the ship's being upstream and B would designate the ship's being downstream. So (1) says that if S perceives an event, such as the ship moving from an upstream to a downstream position, then S's perceptions of the stages of the event must occur in the same temporal order as did the stages of the event: they are “irreversible.” I have placed the term (modal operator) “necessarily” in front of the statement, so that it applies to (or “governs”) the if–then relation expressed by the statement as a whole, in order to bring out a claim made by Strawson (1966: 136) which seems correct. This is that (1) is a conceptual or analytic truth. Strawson bases this claim on the two more basic claims, which also seem correct, that:

- (a) It is a conceptual truth about sense perception that our perceptions of an object are caused by that very object.
 (b) It is a conceptual truth about causation that an effect cannot precede its cause, but must occur at the same time as or after its cause.¹⁰

He notes that (1) follows from these two conceptual truths, and so is itself a conceptual truth, provided one stipulation is made. The stipulation is that there must not be any difference in the causal conditions of the two perceptions as a result of which the perception of *A* occurs after the perception of *B*. For otherwise, one can think of cases where the earlier stage of an event is perceived after the later stage. For example, one could see the ship in its upstream position after seeing it in its downstream position, if the light from its upstream position were delayed by being reflected back and forth several times between mirrors; or one might *hear* a whistle blast that the ship emitted upstream after *seeing* the ship downstream, simply because sound travels slower than light. (Notice that in such cases the conceptual truth that an effect cannot precede its cause is not violated, because both perceptions still occur after their own causes.) Strawson points out that such cases can be circumvented by stipulating that the perceptions of *A* and *B* must be equally direct and in the same sensory mode, or by stipulating that there can be no difference in the causal conditions of the perceptions such that the perception of *A* occurs after the perception of *B*. Provided such a stipulation is understood, Strawson seems right to maintain that (1) is an analytic or conceptual truth.

The second stage of the irreversibility argument moves from (1) to the conclusion that the causal principle is true. In terms of the formulation being constructed here, this is to say that it moves from (1) to the conclusion: (C) Necessarily, if *A* occurs, then *B* occurs. Here the point of the modal term "necessarily" is just to say that the transition from *A* to *B* is governed by whatever type of "necessity" characterizes causation; or, as Kant puts it, that *B* follows upon *A* "according to a rule." Of course, if the Humean regularity view of causation discussed in Chapter 4 is correct, then the "necessity" or "rule" in question reduces to a contingent but exceptionless regularity. But, for the moment, we shall talk heuristically in terms of causal necessity; in due course we shall see how what is said below obtains even when we think of causality purely in terms of regularity or of Hume's "constant conjunction."

It is obvious that (C) does not logically follow from (1) alone; another premiss is needed for (C) to be entailed. This premiss can only be:

- (2) *If necessarily when S perceives an event A-B, then S's perceptions occur in the order A, B, then necessarily if A occurs, B occurs.*

Now if the argument from (1) and (2) to (C) is really Kant's argument, then Strawson is certainly right to say that Kant has committed "a *non-sequitur* of numbing grossness" (1966: 137). The problem is not that the argument as we have formulated it is logically invalid; for its form is:

- (1) Necessarily ($P \supset Q$)
 (2) [$\text{Necessarily } (P \supset Q)$] \supset necessarily ($R \supset S$)

 \therefore (C) Necessarily ($R \supset S$)

which is a perfectly valid (modal) *Modus Ponens*. The “numbing non-sequitur” occurs, rather, within (2), in the transition from its complex antecedent to its consequent. For the antecedent says that when we are perceiving an event, our perceptions of the stages of the event necessarily occur in the same order as those stages themselves. But the consequent says that the stages of an event necessarily occur in a certain order. In other words, the premiss as a whole says that if the order of our perceptions of an event necessarily corresponds to the order of the stages of the event, then the stages of the event necessarily have a certain order. This is fallacious; for it involves, as Strawson shows, a double equivocation on the notion of necessity.

- 1 The *sense* of necessity is not the same in the antecedent as in the consequent. In the antecedent necessity refers to conceptual or analytic necessity, as we explained earlier; while in the consequent it refers to causal necessity (however that kind of necessity is understood).
- 2 The *application* of the notion of necessity is not the same in the antecedent as in the consequent. In the antecedent, what is asserted to be necessary is the correspondence between the temporal order of the stages of an event and the temporal order of our perceptions of those stages; while in the consequent the notion of necessity is applied to the relation between the stages of an event themselves.

Strawson aptly sums up the situation: “It is a very curious contortion indeed whereby a conceptual necessity based on the fact of a change is equated with the causal necessity of that very change” (1996: 138).

The fallacy can be brought out, as we said above, even if we think of causality purely in terms of regular succession rather than causal necessity. To assert that an event $A-B$ has a cause is, then, to assert that there is a kind of event, E , such that events of kind E are regularly followed by events of the kind to which $A-B$ belongs. Keeping this in mind, what (2) asserts is that if our perceptions of the sequence $A-B$ must occur in the order A, B , then there is a kind of event E such that all events of that kind are followed by events of the kind to which $A-B$ belongs. Clearly this is fallacious: from the fact that our perceptions of a sequence $A-B$ must occur in the same order as the members of that sequence, one cannot conclude that there is a kind of event E such that sequences like $A-B$ invariably follow upon events of kind E .

We have dwelt on the irreversibility argument at some length, because it is quite commonly thought to reflect Kant's own thinking in the second analogy. Since it is a fallacious argument, however, and because Kant's second analogy is a difficult text that lends itself to more than one interpretation, it is natural to wonder whether Kant has a better argument for the causal principle. Kant scholars have offered many reconstructions of his reasoning, and these in turn have been criticized by other scholars. We cannot survey this ongoing debate here. Rather, I shall conclude this chapter by presenting one reading of Kant's argument, from Paul

Guyer's book *Kant and the Claims of Knowledge* (1987: Chapter 10), which seems more faithful to Kant's text as well as more promising than others.¹¹ Guyer's discussion is complex and richly ramified; we shall present only the core of his interpretation.

The question Kant is raising in the second analogy could be put in this way: how can I know by observation that event E is occurring? If we think of an event as Kant does, as a transition from a state A to a state B, then this question can be put also as follows: how can I know by observation that a state A is followed in time by a state B? Now Kant's key point, that one's perceptions are successive or serial regardless of whether one is perceiving an event or an enduring state of affairs, means that I cannot know that A is followed in time by B just by knowing that my perception of A is followed by my perception of B, since perceptions of co-existing states of an enduring object would also occur successively, as happens when one views the different sides of a house. According to the irreversibility criterion, the way in which I am supposedly able to tell that A is followed by B in time is by knowing that my perceptions of A and B are irreversible – that they could not have occurred in the order B, A, rather than A, B. However, as Guyer rightly notes (1987: 256), I *cannot* really tell that A is followed in time by B by knowing that my perceptions of A and B are irreversible. For I can know that they are irreversible only if I *already* know that A and B are occurring in the order A, B. This relatively simple point seems to me to be Guyer's key insight. For, in the first place, it goes directly against the "irreversibility" reading of Kant's argument, by showing that reversibility–irreversibility could not really be the criteria we use for determining perceptually whether we are observing an event or an enduring state of affairs. But, even more importantly, Guyer's point seems to be just what Kant needs in order to make his argument work. For if I cannot tell that I am perceiving that A is followed by B either by knowing my perceptions of A and B are successive or by knowing they are irreversible, then how can I tell this? The only answer available seems to be: by knowing that state B follows state A according to a rule; that is, that the event constituted by the transition from A to B has some *cause*. Note also that the irreversibility of my perceptions of A and B is a *consequence* of the fact that B follows A according to a rule, rather than a criterion for deciding whether A was followed in time by B.

To convey the power of this interpretation, I quote two passages: first the passage from Kant that seems to best support it, and then a fairly long passage from Guyer that contains the core of his interpretation. The passage from Kant is this:

Let us suppose that there is nothing antecedent to an event, upon which it must follow according to a rule. All succession of perception would then be only in the apprehension, that is, would be merely subjective, and would never enable us to determine objectively which perceptions are those that really precede and which are those that follow. We would then have only a play of representations, relating to no object I could not then assert that the two states follow upon one

another in the field of appearance [by “field of appearance,” Kant here means the objects perceived, such as the moving ship or the house], but only that one apprehension follows upon the other

If, then, we experience that something happens, we in so doing always presuppose that something precedes it, on which it follows according to a rule. Otherwise I should not say of the object that it follows. For mere succession in my apprehension, if there be no rule determining the succession in relation to something that precedes, does not justify me in assuming any succession in the object. I render my subjective synthesis of apprehension objective only by reference to a rule in accordance with which the appearances [again, “appearances” here means the objects or events perceived], in their succession, that is, as they happen, are determined by the preceding state. The experience of an event (i.e. of anything as *happening*) is itself possible only on this assumption.

(1963: A194–5/B239–40)

And here is the passage from Guyer:

[T]he present problem is only that of distinguishing between an event occurring among represented states of affairs from the event of a change in representations [i.e. perceptions] themselves Thus, the significance of the irreversibility of a sequence of representations . . . is only that such a fact would be a *consequence* of the occurrence of an event in what is being perceived, which *could* be used as a *symptom* of the occurrence of an event *if* it were directly given to consciousness. But what Kant's underlying assumption means is precisely that such a modal fact about the sequence of perceptions is *not* given to consciousness by apprehension alone So Kant's idea is that no alternative remains but that the occurrence of an event be inferred by *adding* to the omnipresent succession of mere representations a *rule* from which it can be inferred that in the circumstances at hand *one state of affairs could only* succeed the other, and *therefore* also that one *representation* could only succeed the other Only from a rule which says that one of the represented states *must* succeed the other can it be inferred that it *does* succeed the other. For . . . though their succession *could* be inferred from the *necessary* sequence or irreversibility of the representations of them if such irreversibility *were* [directly given to consciousness] – since the irreversibility of their representations would be a genuine consequence of the represented states of affairs – the necessity of the sequence of representations is . . . *not* directly given to consciousness. So nothing remains but to invoke a rule from which it follows that one objective state can only succeed and not coexist with the other, from which it *also* follows . . . that the *representation* of the one state not only does but

also only could succeed the representation of the other And a rule which dictates that in a given situation one state of affairs must succeed another is just what Kant means by a causal law. Thus, judgments that events occur are possible only if the states of affairs which comprise them are linked by causal laws.

(Guyer 1987: 247–9)

The argument that Guyer finds in Kant's text may be summarized in this way:

- (1) We cannot know by observation that an event – a transition from a state A to a state B – is occurring by knowing that the perceptions of A and B occur in the order A, B, or by knowing that the perceptions of A and B are irreversible.¹²
- (2) If (1), then the only way we can know by perception that an event – a transition from a state A to a state B – is occurring, is by knowing that B follows A according to a rule; which is to say that the event has a cause.
- (3) If the only way we can know by perception that an event – a transition from a state A to a state B – is occurring, is by knowing that B follows A according to a rule (that the event has a cause), then any event such that we can know of its occurrence by perception must have a cause.

∴ (4) Any event such that we can know of its occurrence by perception must have a cause.

Notice that the conclusion of this argument does not mean that we can know by perception that an event is occurring only if *every* event has a cause. Rather, it means that we can know by perception that an event is occurring only if *that* event has a cause. Notice also that, even if this argument is sound, it does not show that it would be impossible for the world we inhabit not to be governed by causal laws, or impossible for events not to have causes. Rather, it shows only that in such a world we would be unable to know by perception that events were occurring. Thus, Kant's transcendental argument for the causal principle, even if it is sound, constitutes a qualified answer to Hume's critique of that principle, showing, to quote William Brenner (1989: 128) once again, that "the principle of causality is *presupposed* by the perception of events, not *derived from* it."¹³

6

THE BELIEF IN THE EXISTENCE OF BODY

1 Introduction

The purpose of this chapter is to present and discuss Hume's views about our belief in the existence of material things, especially as he presents them in a famous section of the *Treatise* called "Of Scepticism with regard to the Senses." Concerning this section, Jonathan Bennett has written:

It is extremely difficult, full of mistakes, and – taken as a whole – a total failure; yet its depth and scope and disciplined complexity make it one of the most instructive arguments in modern philosophy. One philosopher might be judged superior to another because he achieved something of which the other was altogether intellectually incapable. By that criterion Hume surpasses Locke and Berkeley – because, and only because, of this one section.

(Bennett 1971: 313)

Taking Bennett's words as a cue, the discussion will try to clear up some of the difficulties one encounters when reading this section of the *Treatise*, indicating where I believe Hume makes mistakes, and trying to identify ways in which his discussion is nonetheless instructive.

2 Three assumptions behind Hume's account

Hume's entire account of what he calls the "belief in the existence of body" rests on three assumptions. In calling these "assumptions," I do not mean to say that Hume never gives reasons for them, but rather that the emphasis of his discussion falls more heavily on developing their implications than it does on justifying them – more on working *from* them as "givens" than on working *to* them as points to be established. In this part, accordingly, we shall expound the three assumptions without criticizing them; though in part 6 we shall critically discuss the first assumption.

The first and most basic assumption behind Hume's account of our belief in the existence

of body is one which was a commonplace of seventeenth- and eighteenth-century philosophy and still has some defenders today. This is the view Berkeley expressed by saying that the things we perceive by our senses are nothing but our own “sensations or ideas.” Put in Hume’s terminology, this view is that the only things we perceive by our senses are our own perceptions, more specifically, our own impressions (of sensation). As can be seen from the following passages, Hume regards this view as virtually axiomatic:

We may observe, that ’tis universally allowed by philosophers, and is besides pretty obvious of itself, that nothing is ever really present with the mind but its perceptions or impressions and ideas . . . nothing is ever present to the mind but perceptions . . .

(T:67)

[P]hilosophy informs us, that every thing, which appears to the mind, is nothing but a perception, and is . . . dependent on the mind . . .

(T:193)

[N]othing is ever really present to the mind, besides its own perceptions

(T:197)

[T]he slightest philosophy . . . teaches us, that nothing can ever be present to the mind but an image or perception, and that the senses are only the inlets, through which these images are conveyed, without being able to produce any immediate intercourse between the mind and the object.

(E:152; S:104; F:183)

The most vulgar philosophy informs us, that no external object can make itself known to the mind immediately, and without the interposition of an image or perception. That table, which just now appears to me, is only a perception, and all its qualities are qualities of a perception.

(T:239)

The mind has never anything present to it but the perceptions, and cannot possibly reach any experience of their connexion with objects.

(E:153; S:105; F:184)

This *first* key assumption may strike you as strange, especially if Hume is one of the first philosophers you are studying. It will help, therefore, to begin by describing in summary

fashion the motivation for the assumption. Since our purpose here is to understand where Hume is “coming from” – to see things from his point of view – I will for the time being simply expound the motivation behind his key assumption, postponing critical discussion of it until part 6.

The grounds for Hume's assumption consist chiefly in – what is often called – the “argument from illusion.” This label is something of a misnomer, since it covers a number of arguments, not all of which pertain to “illusions” in any ordinary sense of the term. Let us briefly consider some of these arguments.

The first is the *argument from perceptual relativity*. This argument is itself really a family of arguments which have been known at least since Plato's day and whose offspring are legion in the philosophy of perception. Such arguments appeal to the consideration that the way things appear to us in perception depends on a host of conditions, other than just those things' properties, in order to support various philosophical conclusions about perception and its objects. Plato, for example, used arguments, particularly in the *Theatetus* (see 151D–186E), from perceptual relativity to show that perception is not knowledge; and Berkeley used them in his *Three Dialogues* to show that sensible qualities are nothing but sensations or ideas (Berkeley 1993a: 168–82 especially). In the version most directly relevant to Hume's first assumption, the argument is supposed to show that what we really perceive by our senses is not bodies themselves, but rather our own impressions, “perceptions,” or “images.” Hume himself offers the following brief statement of the argument:

The table, which we see, seems to diminish, as we remove further from it: but the real table, which exists independent of us, suffers no alteration: it was, therefore, nothing but its image, which was present to the mind.

(E:152; S:104; F:183)

In other words, when one looks at an ordinary object like a table, what one sees changes as one's distance from the table and one's angle of vision change. But the table itself does not change. So, Hume concludes, what one sees is not really the table, but only a visual image, or rather a series of different images – impressions of sense, “sensations or ideas;” or, to use the twentieth-century term, “sense-data.”

A second argument is the *argument from the causal facts of perception*. When one's sense-receptors (one's eyes, ears, nose, etc.) are stimulated by an ordinary object, what one perceives depends causally on the sense-receptors one possesses and on the condition (i.e. functional capability) of those organs. For example, when the eyes of a human being and the eyes of a housefly are each stimulated by a lump of sugar, what the human and the fly respectively see is presumably very different, because the eyes of a human and the eyes of a fly are very different. Or when the eyes of a near-sighted person, those of a person with 20–20 vision, and those of a far-sighted person are stimulated by a United States' flag, what

each sees is different, because of the different capabilities (condition) of their eyes. But the properties of the stimulus-objects – the lump of sugar and the flag – do not causally depend on the type or the condition of anyone’s sense-receptors; they are utterly independent of such facts. So, again, it is inferred that what one sees is not the stimulus-object itself, but rather an image or percept the qualities of which depend, at least in part, on the perceiver’s sense-receptors.

A third argument is the *argument from illusion* (where the term “illusion” now refers to genuine cases of illusion). Hume gives a version of this argument, in a passage that also adduces perceptual relativity and the causal facts of perception:

’Twill . . . be proper to observe a few of those experiments, which convince us, that our perceptions are not possest of any independent existence. When we press one eye with a finger, we immediately perceive all the objects to become double, and one half of them to be remov’d from their common and natural position. But as we do not attribute a continu’d existence to both these perceptions, and as they are both of the same nature, we clearly perceive, that all our perceptions are dependent on our organs, and the disposition of our nerves and animal spirits. This opinion is confirm’d by the seeming encrease and diminution of objects, according to their distance; by the apparent alterations in their figure; by the changes in their colour and other qualities from our sickness and distempers; and by an infinite number of other experiments of the same kind; from all which we learn, that our sensible perceptions are not possest of any distinct or independent existence.

(T:210–11)

In other words, by pressing one eye with a finger, one can cause oneself to see everything double. But in such a case, at least half of the things one sees are merely mental images that depend for their very existence on being perceived. Furthermore, there is no special, tell-tale, qualitative difference between these images and the other half of the things seen to indicate that while the former are merely mental images, the latter are material things that exist whether or not they are being perceived. So it is inferred that all the things seen are just mental images. A rather similar argument, called the “argument from hallucination,” can be constructed by appealing to the occurrence of hallucinations. Since, again, there is no special discernible difference between what is experienced in a vivid hallucination and in a case of normal perception to indicate that only in the hallucination do we perceive merely a mental image, it is inferred that we perceive merely a mental image in both cases.

Fourth, and finally, *epistemological arguments* also have provided an impetus for the view that we perceive only our own impressions or “sense-data.” One such argument – perhaps the strongest – goes as follows. Whatever else perception may be, we take it to be a way of acquiring knowledge of its objects. We firmly believe that by perceiving an object,

one can come to know both that it exists and that it has certain properties. However, there is good reason to think that any perceptual experience had when a physical object is stimulating one's sense-receptors can be exactly duplicated when there is no such object stimulating one's sense-receptors. Such a duplicate experience can be produced, for example, by directly stimulating the brain, or in a drug-induced hallucination, or in a naturally occurring hallucination, or in a vivid dream. Conceivably, it could even be produced by a deceiving God, or by an evil deceiver like the one postulated by Descartes.¹ Therefore, it is inferred, the objects whose existence and properties we come to know in perception cannot really be physical objects that stimulate our sense-receptors, since sensory experiences indistinguishable from those they would thereby cause could be caused also in these other ways. What then are these objects? The answer given is that they are our own impressions or sense-data – entities that we experience regardless of the way the experience is generated. Only so, it seems, can we defend our deeply held conviction that in sense perception we acquire knowledge of the existence, and of at least some of the properties, of the objects perceived. Of course, however, the impressions or sense-data whose existence and nature we come to know are themselves not physical objects. Rather, it is claimed, our knowledge of physical objects must be somehow inferred or derived from the impressions or sense-data.

This last point brings us to the *second* assumption behind Hume's account – one that he shares with Berkeley, though this time not with most other seventeenth- and eighteenth-century philosophers. This assumption is that from knowledge of our impressions, we cannot legitimately infer the existence of bodies existing outside our minds. Many philosophers of the Modern period, including notably Descartes and Locke, have held that our knowledge of the existence of material things – of the entire material world – is based on an inference or argument from the occurrence of our sensations, ideas, or impressions to the existence of material things. This inference, though construed quite differently by individual philosophers, is basically a causal one, going from effects (the sensations, ideas, or impressions) to their supposed causes (material things, bodies).² The inference is supposed to show, not only that the impressions are caused by bodies, but also that they resemble those bodies to a certain extent; so that they provide knowledge of bodies by being representations of those bodies. The overall position that results – we perceive only our own impressions (or sensations, or ideas), but can infer from them the existence of bodies which cause them and which they represent – is commonly called “the causal theory of perception” or “the representational theory of perception.”

Berkeley powerfully criticized this theory by challenging its proponents to show why our impressions could not be produced, as Descartes himself had initially wondered, in some quite different way. Thus, Berkeley wrote:

[W]hat reason can induce us to believe the existence of bodies without the mind, from what we perceive . . . ? I say it is granted on all hands (and what happens in dreams, phrensies, and the like, puts it beyond dispute) that it is possible we might be affected with all the ideas we have now, though no bodies existed without,

resembling them. Hence it is evident the supposition of external bodies is not necessary for the producing our ideas: since it is granted they are produced sometimes, and might possibly be produced always in the same order we see them in at present, without their concurrence Suppose, what no one can deny possible, an intelligence, without the help of external bodies, to be affected with the same train of sensations or ideas that you are, imprinted in the same order and with like vividness in his mind. I ask whether that intelligence hath not all the reason to believe the existence of corporeal substances, represented by his ideas, and exciting them in his mind, that you can possibly have for believing the same thing? Of this there can be no question; which one consideration is enough to make any reasonable person suspect the strength of whatever arguments he may think himself to have, for the existence of bodies without the mind.

(Berkeley 1993b: 95–6)

Berkeley, then, strongly maintains that the causal-representational theory of perception unwittingly leads to *scepticism*: since what it proposes as the only way to arrive at knowledge of bodies is an indefensible causal inference.

Hume fully endorses this Berkeleyan critique of the theory; indeed, he supplements it with his own original argument against the theory. This influential argument, which Hume gives in both the *Treatise* and the *Enquiry*, goes as follows:

The only existences, of which we are certain, are perceptions, which being immediately present to us by consciousness, command our strongest assent, and are the first foundation of all our conclusions. The only conclusion we can draw from the existence of one thing to that of another, is by means of the relation of cause and effect, which shews, that there is a connexion betwixt them, and that the existence of one is dependent on that of the other. The idea of this relation is deriv'd from past experience, by which we find, that two beings are constantly conjoin'd together, and are always present at once to the mind. But as no beings are ever present to the mind but perceptions; it follows that we may observe a conjunction or a relation of cause and effect between different perceptions, but can never observe it between perceptions and objects. 'Tis impossible, therefore, that from the existence or any of the qualities of the former, we can ever form any conclusion concerning the existence of the latter, or ever satisfy our reason in this particular.

(T:212)

It is a question of fact, whether the perceptions of the senses be produced by external objects, resembling them: how shall this question be determined? By experience surely; as all other questions of a like nature. But here experience is, and

must be entirely silent. The mind has never anything present to it but the perceptions, and cannot possibly reach any experience of their connexion with objects. The supposition of such a connexion is, therefore, without any foundation in reasoning.

(E:153; S:105; F:184)

This argument stems directly from Hume's analysis of causal reasoning; it shows that when that analysis is combined with the view that we perceive only our own impressions, the result is that a causal inference from the impressions to the existence of bodies is completely worthless. Of course, Hume is committed by his critique of induction to the view that causal inferences cannot be rationally justified. Nevertheless, it is clear from many places in his writings that he regards reasoning from past experience as being more legitimate than, say, superstition or mere guesswork. Indeed, in the *Treatise*, he even includes a chapter, entitled "Rules by which to judge of causes and effects," offering eight rules that are hard to interpret otherwise than as criteria for distinguishing between "better" and "worse" causal inferences (T:Book I, part iii, Section 15). To what extent this is consistent with Hume's critique of induction, and to what extent it argues for a more "naturalistic" reading of Hume that de-emphasizes the negative aspect of his position, are controversial questions of interpretation that we have not tried to resolve (although we have argued that Hume's thesis according to which inductive inferences cannot be rationally justified does not commit him to the view that such inferences are not rational).³ But the point to be made here is that, in the passages just quoted, Hume is arguing that the causal inference from impressions to objects, quite apart from what may be said about other causal inferences or causal reasoning in general, is *wholly worthless*: even if other causal inferences, such as those from past to future constant conjunctions, could be given a rational justification, the inference from impressions to objects would still be illegitimate. The reason why this is so can be stated as a simple argument: We can establish a causal relation between *As* and *Bs* only by observing that *As* and *Bs* have been constantly conjoined.

- (1) We can establish a causal relation between *As* and *Bs* only by observing that *As* and *Bs* have been constantly conjoined.
- (2) We can observe that *As* and *Bs* have been constantly conjoined only if we can perceive *As* and we can perceive *Bs*.
- (3) We can perceive impressions but we cannot perceive bodies.

∴ We cannot establish a causal relation between impressions and bodies.

The *third* assumption behind Hume's account is his rejection of Berkeley's positive theory of perception and its objects. Berkeley argued for the novel view, now known as "immaterialism," that matter does not exist, and that what people call "material objects" or "bodies" are really nothing but ordered groups of sensations or ideas. As he puts it:

Thus, for example, a certain colour, taste, smell, figure and consistence having been observed to go together, are accounted one distinct thing, signified by the name *apple*. Other collections of ideas constitute a stone, a tree, a book, and the like sensible things.

(Berkeley 1993b: 89)

One chief benefit of this theory, as Berkeley sees it, is that it eliminates the need for the indefensible causal inference from ideas to bodies, since bodies are themselves only groups of ideas. Thus, by reducing bodies to ideas, Berkeley takes himself to have disposed of scepticism about our knowledge of bodies; he sees his immaterialism as the best cure for scepticism. Hume, however, rejects Berkeley's reduction of bodies to ideas. His general attitude to Berkeley's views is encapsulated in the following passage:

[M]ost of the writings of that very ingenious author form the best lessons of scepticism, which are to be found among the ancient or modern philosophers, Bayle not excepted. He professes . . . to have composed his book against sceptics. . . . But that all his arguments, though otherwise intended, are, in reality, merely sceptical, appears from this, *that they admit of no answer and produce no conviction*.

(E:155n; S:106–7n 64; F:186)

It should now be clear that Hume is committed to a deep scepticism with respect to our knowledge of a material world. For he assumes that

- (1) we perceive only our own impressions;
- (2) any inference from those impressions to the existence of bodies causing the impressions is wholly illegitimate; and
- (3) bodies are not merely groups of impressions.

These three assumptions imply that our belief in the existence of bodies is completely unfounded. Everything Hume says about this belief – and he has much more to say about it – needs to be understood against this sceptical background.

3 The general nature of Hume's account

Although Hume holds that our belief in the existence of bodies has no rational foundation, he holds also that humans have an irresistible propensity to believe that bodies exist. Indeed, this propensity is so powerful that, most of the time, we have no choice but to yield to it. It is only while we are actually engaged in philosophical reflection on the topic of sense perception that we can bring ourselves to suspend judgment regarding the existence of

bodies. Even the sceptic, as soon as he or she ceases to focus on the arguments that show this belief to be without any rational foundation, falls back into the notion that the very things he or she perceives *are* bodies. This contrast between the results of reasoned reflection, on the one hand, and what we spontaneously and naturally believe, on the other hand, is part and parcel of Hume's naturalism, which has been so strongly emphasized by recent Hume scholars. In the section of the *Treatise* that we are examining (Book I, part iv, section 2), he expresses the contrast in these words:

There is a great difference betwixt such opinions as we form after a calm and profound reflection, and such as we embrace by a kind of instinct or natural impulse, on account of their suitableness and conformity to the mind. If these opinions become contrary, 'tis not difficult to foresee which of them will have the advantage. As long as our attention is bent on the subject, the philosophical and study'd principle may prevail; but the moment we relax our thoughts, nature will display herself, and draw us back to our former opinion.

(T:214)

So, near the end of the section, after going through a series of arguments designed to show that our belief in the existence of bodies is shot through with errors and has no rational justification, he concludes with these ironical words: "I . . . take it for granted, whatever may be the reader's opinion at this present moment, that an hour hence he will be persuaded there is both an external and internal world" (T:218).

Another way of putting Hume's point – that the conclusions of philosophical reflection are inevitably overshadowed by our natural propensity to believe in external objects – is this: although our belief in the existence of bodies cannot be the result of rational reflection, it is the product of certain psychological principles of human nature. In other words, the belief cannot be rationally justified, but it can be psychologically or naturalistically explained by an adequate "science of man." This is of course Hume's position, and it explains why he introduces the section as he does: "We may well ask, *What causes induce us to believe in the existence of body?* but 'tis vain to ask, *Whether there be body or not?* That is a point, which we must take for granted in all our reasonings" (T:187). Hume's point is that philosophy wastes its time when it asks whether bodies exist or not, because no amount of argument or reasoning can show that they exist, and our human nature compels us anyway to believe that they exist. Rather, the proper business of philosophy is to ask: what are the principles of human nature that make us believe that bodies exist? Accordingly, Hume announces that "the subject, then, of our present enquiry is concerning the *causes* which induce us to believe in the existence of body" (T: 187–8).

Notice that this approach to the belief in body is closely parallel to Hume's treatment of induction and of personal identity. When Hume finds that induction cannot be rationally justified, he is not content with this negative conclusion. Rather, he goes on to give a psychological explanation – in terms of his principle of association of ideas – of inductive

inferences. This explanation also provides the material for his account of the idea of necessary connection, as a projection into the phenomena we perceive of our feelings of expectation. As for personal identity, recall that Hume finds there is no impression of the self from which can be derived any idea of a continuing self persisting through time. Coupled with his strict view of identity as requiring no change and no interruption, this finding leads him to the conclusion that the self possesses no real identity through time. But again, Hume is not content to stop with this negative conclusion; instead, he goes on to offer an explanation, in terms of psychological principles, of why we nevertheless ascribe to ourselves identity through time and change. Thus, Hume takes a similar approach with respect to induction, the belief in personal identity, and the belief in the existence of body. First he argues that these embody various errors and have no rational foundation; then he gives a psychological explanation of them. This approach, which is especially prominent in the *Treatise* but by no means absent from the *Enquiry*, reflects Hume's desire to explode rationalist ways of thinking and to replace them with his naturalistic science of human nature.

4 The nature and origin of the belief in the existence of body

Hume begins his inquiry into the causes of our belief in body by analyzing or breaking down this belief into two parts:

- (1) the *continued* existence of the objects of the senses while they are not being perceived; and
- (2) the existence of the objects of the senses *distinct* from (independently of) being perceived.

Hume means that our belief in the existence of bodies boils down to a belief in (1) and (2); that is to say, it consists in believing that the things we perceive by our senses *continue* to exist at times when we are not perceiving them, and that they exist *independently* of whether or not we are perceiving them. Thus, for example, your belief that Mount Everest exists consists in your believing that Mount Everest continues to exist when neither you nor anyone else is perceiving it, and that it exists regardless of whether anyone is perceiving it. Hume says that (1) and (2) are logically related. For if (1) is true – if sensible objects continue to exist while not being perceived – then (2) must be true also; that is, sensible objects must exist independently of being perceived. Hume also says that, conversely, if (2) is true, then (1) is true: he thinks that (1) and (2) are mutually entailing beliefs, and so stand or fall together. This, however, is questionable; for why could there not be momentary objects which, though they do not depend on being perceived for their existence, happen to exist only when they are perceived? If there could be such objects, then Hume is wrong to think that (2) – distinct or independent existence – entails (1) – continued existence. But this matters little to his discussion, which, as we shall see, requires only that (1) entail (2). Hume

himself emphasizes this entailment when he says that “the opinion of the *continu'd* existence of body . . . is prior to that of its *distinct* existence, and produces that latter principle” (T:199), and that “’tis the opinion of a *continu'd* existence, which first takes place, and without much study or reflection draws the other along with it, wherever the mind follows its first and most natural tendency” (T:210).

Hume, nevertheless, devotes some space, at the beginning of his discussion, to the belief in the independent or “distinct” existence of the objects we perceive. He first offers some arguments to show that this belief cannot arise from the *senses* themselves; let us look at two of these arguments.

The first argument is that our impressions themselves have no distinct existence; so the only way our senses could produce a belief in distinct existence would be if we perceived those impressions as images or representations of some other distinct objects. But, Hume says:

That our senses offer not their impressions as the images of something *distinct*, or *independent*, and *external*, is evident; because they convey to us nothing but a single perception, and never give us the least intimation of any thing beyond. A single perception can never produce the idea of a double existence.

(T:189)

The second argument is designed to defeat a possible explanation of how our impressions give rise to the belief in distinct existence, despite the fact that they are not perceived as images of distinct objects. This possible explanation is that the belief in distinct existence arises from the fact that we perceive impressions as external to our own bodies and hence as external to ourselves.⁴ As Hume puts it:

[O]ur own body evidently belongs to us; and as several impressions appear exterior to the body, we suppose them also exterior to ourselves. The paper, on which I write at present, is beyond my hand. The table is beyond the paper. The walls of my chamber are beyond the table. And in casting my eye towards the window, I perceive a great extent of fields and buildings beyond my chamber. From all this it may be infer'd, that no other faculty is requir'd, besides the senses, to convince us of the external existence of body.

(T:190–1)

Hume rejects this explanation, for the following reason (among others):

[P]roperly speaking, ’tis not our body we perceive, when we regard our limbs and members, but certain impressions, which enter by the senses, so that the ascribing

a real and corporeal existence to these impressions, or to their objects, is an act of mind as difficult to explain, as that which we examine at present.

(T:191)

In other words: since we do not really perceive our own bodies but only certain impressions of our own arms, hands, fingers, etc., the fact that we have impressions also of other things – of a paper, table, walls, and fields – does not amount to perceiving those items to be exterior to our bodies or to ourselves; it amounts only to perceiving an array of impressions that includes the impressions both of our own bodily parts and those of other things.⁵ The explanation proposed of why we believe the latter to be exterior to ourselves could work only if we already had an explanation of why we believe the former to have a “distinct” existence – which is the very sort of explanation we are looking for and have not yet found.

Having argued that the belief in the independent or “distinct” existence of the objects of sense perception cannot arise from the senses themselves, Hume argues next that it cannot be based on reason. His argument is contained in the following passage:

[P]hilosophy informs us, that every thing, which appears to the mind, is nothing but a perception, and is interrupted, and dependent on the mind; whereas the vulgar confound perceptions and objects; and attribute a distinct continu'd existence to the very things they feel and see. This sentiment, then, as it is entirely unreasonable, must proceed from some other faculty than the understanding. To which we may add, that as long as we take our perceptions and objects to be the same, we can never infer the existence of the one from that of the other, nor form any argument from the relation of cause and effect; which is the only one that can assure us of matter of fact. Even after we distinguish our perceptions from objects, 'twill appear presently, that we are still incapable of reasoning from the existence of the one to that of the other: So that upon the whole our reason neither does, nor is it possible it ever shou'd, upon any supposition, give us an assurance of the continu'd and distinct existence of body.

(T:193)

Here Hume does not separate the issue of “distinct” existence from that of “continu'd” existence; he treats the two together. But insofar as the argument bears just on “distinct” existence, it seems to boil down to the simple reasoning that since it is so easily shown to be *false* that the objects of sense perception have such an existence – since the slightest philosophy “informs us” that they are “dependent on the mind” – it follows that any belief to the contrary must be “entirely unreasonable,” and so not based on reason. To be sure, Hume also says that we cannot *both* identify perceptions and objects *and* regard the latter as causes of the former, but this point does nothing to support the claim that belief in distinct existence cannot be based on reason; rather it has force only against someone who would say

that impressions are both the sole objects of perception and that they are caused by objects of perception – an absurd position that no philosopher would maintain. Hume also says that once we distinguish perceptions from objects, the causal inference from the perceptions to the existence of objects is illegitimate; but this point, though it foreshadows his argument against causal theories of perception, has no bearing on the question of whether the belief that *the very objects we perceive* have a “distinct” existence is based on reason. So Hume’s argument that this belief is not based on reason does seem to reduce to saying that the belief is so easily shown to be false that it cannot be reasonable.⁶

Hume, then, thinks he has shown that our belief in the “distinct” or independent existence of the objects of sense perception cannot arise either from the senses themselves or from reason. But what about our belief in the “continu’d” existence of these objects: what is its status, and how does it arise? This question brings us to the heart of Hume’s discussion, most of which is directly addressed to explaining the belief that the objects of sense perception continue to exist when they are no longer being perceived. Further, if Hume can explain this belief, then he will also have explained the belief that these objects have a “distinct” existence – that they exist independently of perception – since, as we have seen, the former belief entails the latter.

For brevity’s sake, we shall henceforth often call the belief that the objects of sense perception continue to exist when they are no longer being perceived the “belief in object-continuity.” (Note, then, that the term “object,” when it occurs within this phrase, will always refer to the very object which is perceived by the senses.) Now the central theme of Hume’s discussion of our belief in object-continuity is that this belief is simply false, and that our firm acceptance of this falsehood is precisely what his psychological principles must explain. Let us consider first why Hume thinks the belief is false, and afterwards why he thinks we accept it despite its falsity.

In some places, Hume attacks the belief in object-continuity by attacking the belief in distinctness; he uses a *Modus Tollens* argument that goes this way (see T:210–11, 214):

- (1) If the objects of sense perception continue to exist while not being perceived, then they exist independently of being perceived.
 - (2) The objects of sense perception do not exist independently of being perceived.
- ∴ The objects of sense perception do not continue to exist while not being perceived.

To support this argument’s second premiss, Hume cites the “experiments, which convince us, that our perceptions are not possess of any independent existence,” such as pressing one eye with a finger and seeing everything double – which, as we saw earlier, he regards as showing that we perceive only our own impressions.

Hume has, however, what he regards as a much more fundamental reason against the belief in object-continuity: namely, that this belief is just false on the face of it. This is because our

perceptions are *obviously* discontinuous, “broken,” or “interrupted.” Consider, for example, your present visual impressions of the book you are reading. Those impressions are plainly discontinuous: they vanish each time you glance away from the book, and they cease to exist for the much longer periods of time during which you are asleep, or engaged in activities other than reading this book. This simple and obvious point, Hume thinks, plainly reveals the falsity of the belief in the “continu’d” existence of the objects of sense perception.

To grasp Hume’s thinking here, it is crucial to understand that when he says our perceptions are interrupted or discontinuous, he is not saying merely that our perceivings, or acts of perception, or perceptual episodes, are discontinuous, which is of course true. Rather, he is claiming also that the *objects* of our perceptions – the things that we perceive – are discontinuous. For Hume simply does not distinguish between perceptual episodes and the objects perceived by them; his use of the terms “perceptions” and “impressions” to stand indifferently for both perceptual episodes and their objects is not just linguistic carelessness or sloppiness, but embodies a genuine failure to distinguish between them. The only form in which he recognizes a distinction between perception and its objects is that of the philosophical theory of the “double existence of perception and objects” advocated by Descartes, Locke and others – a theory that Hume strongly rejects. Except within the context of this theory, Hume sees no place for a distinction between perception and its objects.

Hume’s denial of this distinction calls for an explanation, if only because the distinction is so elementary. Even ordinary, unsophisticated, common sense would distinguish between *perceiving* something, or the perception of something, and *what* is perceived: between seeing an apple and the apple which is seen; or between touching a cat and the cat which is touched; or between smelling a flower and the flower which is smelled, etc. Why then does Hume deny this elementary distinction? The reason, I suggest, lies in the basic governing assumption of his account of perception – his thesis that we perceive only our own “perceptions” or impressions. For impressions have exactly the same temporal characteristics, including notably the same duration, as perceptions, taken as perceptual episodes or acts of perceiving. So the thesis that we perceive only impressions implies that the objects of perception last no longer than the perceptions themselves. But in that case, the distinction between perceptions and objects collapses, at least for an Empiricist like Hume. For, as Kant shows, one key difference between perceptions and their objects is precisely their respective temporal properties: on viewing a house, one’s perceptions of the front, sides, and back of the house are successive and may be interrupted, but the parts of the house are co-existent and exist continuously. Furthermore, from Hume’s Empiricist point of view, it seems that this temporal difference is the *only* thing that could differentiate between perceptions and their objects, because any “act” of perception distinct from the object perceived would not be something of which we could have any impression: it would be, so to speak, diaphanous or “transparent.” Thus Hume’s acceptance of the philosophical view that we perceive only our own impressions seems to explain his otherwise puzzling denial of the elementary distinction between

perceptions and objects. By the same token, it explains why it would be so foreign to Hume to distinguish, as Kant does, between the time-relations of perceptual episodes and those of their objects – to say that while the perceptual episodes are always successive, their objects may be co-existent; or that while the perceptions of the front of a house are interrupted during the time that the observer sees the back, and the perceptions of the back are interrupted during the time the observer sees the front, the very front and back of the house that the observer successively sees exist continuously. For Hume, by contrast, the fact that perceptual episodes are discontinuous entails that the objects of perception are discontinuous too.

The chief mistake which Hume accordingly thinks our belief in body involves – that of mistaking manifestly discontinuous items for continuous ones – is so gross that it cries out for an explanation. So, it is this very mistake that Hume's psychological principles are invoked to explain; the bulk of his discussion is an explanation of our (false) belief in the "continu'd" existence of the objects of the senses, or in object-continuity. But since the "continu'd" existence of objects entails their "distinct" existence, Hume thinks, as already mentioned, that if he can successfully explain our belief in the former, he will have explained also our belief in the latter.

Hume asks first whether the belief in object-continuity is based on the senses. He answers that it cannot be:

[T]he SENSES . . . are incapable of giving rise to the notion of the *continu'd* existence of their objects, after they no longer appear to the senses. For that is a contradiction in terms, and supposes that the senses continue to operate, even after they have ceas'd all manner of operation.

(T:188)

In other words, belief that the objects of sense perception continue to exist while they are not being perceived could arise from the senses only if we continued to perceive the objects at the very times when we do *not* perceive them; but we do not continue to perceive them at those times (that would be self-contradictory); therefore, this belief cannot arise from the senses. This argument seems quite decisive.

Hume considers next whether the belief in object-continuity is based on reason. In a passage that we have already quoted, he argues that it is not. Here is the key part of that passage:

[P]hilosophy informs us, that every thing, which appears to the mind, is nothing but a perception, and is interrupted, and dependent on the mind; whereas the vulgar confound perceptions and objects; and attribute a distinct continu'd existence to the very things they feel and see. This sentiment, then, as it is entirely unreasonable, must proceed from some other faculty than the understanding

[U]pon the whole our reason neither does, nor is it possible it ever shou'd, upon any supposition, give us an assurance of the continu'd and distinct existence of body.

(T:193)

As noted earlier, this passage treats “continu’d” and “distinct” existence together; so it is, at least in part, a defense of Hume’s claim that the belief in distinct existence is not based on reason, as we have already seen. But since Hume does not offer any further or separate reasons for denying also that belief in continued existence is based on reason, it seems that he regards the passage as sufficient to support *that* denial as well. As such, it seems to amount to the simple argument that since “every thing, which appears to the mind” is obviously “interrupted,” the “sentiment” that what appears to the mind has a continued existence is therefore “entirely unreasonable,” and so cannot be produced by our reason. Just as in the case of “distinctness,” Hume’s argument seems again to be that just because the belief in continuity is so easily shown to be *false*, it cannot be *reasonable*, and so cannot stem from the use of our reason.

If the belief in object-continuity does not arise from either the senses or the exercise of reason, then how does it arise? Hume’s answer is that our impressions have two features – “constancy” and “coherence” – which cause us mistakenly to believe that the impressions themselves continue to exist between the times that we perceive them. Since this belief is not based on the senses or on reason, and since it amounts to positing something – the continued existence of the impressions – which is in fact untrue, and which is purely a product of our own human invention, Hume ascribes it to the faculty or power of *imagination*. His basic thesis is that the imagination, stimulated by the constancy and coherence of our impressions, naturally “fills in the gaps” between discontinuous impressions, thereby making us believe that they continue to exist between the times we perceive them. As he puts it, in an apt and striking image: “[T]he imagination, when set into any train of thinking, is apt to continue, even when its object fails it, and like a galley put into motion by the oars, carries on its course without any new impulse” (T:198).

To better understand Hume’s basic thesis, we need to see what he means by “constancy” and “coherence.” The reader is warned, however, that Hume’s characterizations of these two features are careless, and need to be corrected in order to serve his purposes. Here is Hume’s characterization of *constancy*:

After a little examination, we shall find, that all those objects, to which we attribute a continu’d existence, have a peculiar *constancy*, which distinguishes them from the impressions, whose existence depends upon our perception. Those mountains, and houses, and trees, which lie at present under my eye, have always appear’d to me in the same order; and when I lose sight of them by shutting my eyes or turning

my head, I soon after find them return upon me without the least alteration. My bed and table, my books and papers, present themselves in the same uniform manner, and change not upon account of any interruption in my seeing or perceiving them.

(T:194–5)

And here is Hume's characterization of coherence:

Bodies often change their positions and qualities, and after a little absence or interruption may become hardly knowable. But here 'tis observable, that even in these changes they preserve a *coherence*, and have a regular dependence on each other; which is the foundation of a kind of reasoning from causation, and produces the opinion of their continu'd existence. When I return to my chamber after an hour's absence, I find not my fire in the same situation, in which I left it: But then I am accusom'd in other instances to see a like alteration produc'd in a like time, whether I am present or absent, near or remote. This coherence, therefore, in their changes is one of the characteristics of external objects, as well as their constancy.

(T:195)

In these passages, Hume's intention is to describe certain kinds of order, uniformity or recurrence in our *impressions*, which he calls "constancy" and "coherence." But instead of confining himself to impressions, he talks about "bodies" – physical things like mountains, houses, trees, books, papers, and his bed, table, and fireplace. Further, he even says that coherence and constancy are "characteristics of external objects" or "bodies," as opposed to features of impressions. This is careless, and Hume should not have done it. For his purpose is to explain, in terms of certain kinds of order or recurrence in our impressions, why we believe that there are physical things or bodies; so his explanation must not appeal to (our belief in) the existence of such things, but only to features of the impressions themselves – a point well made by such commentators as Bennett (1971: 323) and Stroud (1977: 100). One cannot explain *why* it is believed that there are Xs by appealing to the belief *that* there are Xs.

Mindful of this point, let us try to characterize constancy and coherence without assuming the existence of physical things. We may define constancy this way: to say that our impressions have constancy means that our experience presents us with discontinuous sets of impressions whose members exactly resemble each other. For example, on each occasion when Hume has the experience that he would call "looking at my bed and table, books and papers," he has a similar array of visual impressions. This resemblance, Hume thinks, causes the imagination to "fill in the gaps" between these occasions, by supposing that those visual impressions continued to exist during the times he was no longer having such an experience. As for coherence, we may define it this way: to say that our impressions have coherence

means that the temporal relations that we observe between members of an altering but continuous series of impressions are preserved between members of other altering continuous series composed of similar impressions, and also between members of discontinuous series of impressions that resemble members of those continuous series.

Here is an example, based on Hume's illustration. Suppose I have the experience that I would call continuously viewing a slowly diminishing fire in my fireplace. Then, when I have another continuous visual experience whose constituent elements are the same, those elements occur with the same temporal relations as before: I do not, for example, first have impressions of a small fire, then of a medium-sized one, and finally of a large one; rather, I once again have impressions of a large fire, then a medium-sized one, then a small one. Also, when I have an experience that I would call discontinuously viewing such a fire – of seeing the fire in the fireplace, then leaving the area for an hour, and returning to see fire in the fireplace – the relations between the elements of that experience are the same as are the relations between the “matching” elements of the continuous experiences: for example, the impressions of the fire had at the end are impressions of a smaller fire than the impressions of the fire had at the beginning, not of one larger than, or as large as, the fire at the beginning. This coherence among our impressions, Hume thinks, also contributes to the belief in object-continuity; for example, to the belief that when I have impressions of a large fire in my fireplace, followed by no impressions of fire, followed by impressions of a small fire in my fireplace, a slowly-diminishing fire existed unperceived in between.

Hume discusses separately, and in great detail, how constancy and coherence each contribute to the belief in object-continuity. His general position is that coherence alone is “too weak to support so vast an edifice, as is [belief in] the continu'd existence of all external bodies” (T:198–9), and he places the greater weight on constancy. Let us accordingly focus here on constancy; we shall return to coherence in part 7 of this chapter.

To explain the role of constancy, Hume presents a four-part “system” that is supposed to show how the constancy of our impressions gives rise to the belief in their “continu'd existence.” Without trying to cover or discuss all the complexities in his account, its salient points may be summarized in the following way.

The “first part of [Hume's] system” is just his account of identity, which was discussed in Chapter 1. It will be remembered that Hume's account is essentially an account of identity through time: it aims to say what is required for any thing to retain its identity, to remain the same thing, through a period of time. His view, you will recall, is a strict one: identity through time requires not only that there be no break or interruption in a thing's existence throughout the time in question, but also that the thing not change in any way during that time. This definition of identity is stated most explicitly in the section on personal identity, where Hume writes: “We have a distinct idea of an object, that remains invariable and uninterrupted thro' a suppos'd variation of time; and this idea we call that of *identity* or *sameness*” (T:253). But the same notion of identity is operative in the section that we are now discussing, where Hume says:

[T]he *principium individuationis*, or principle of identity . . . [or] principle of individuation is nothing but the *invariableness* and *uninterruptedness* of any object, thro' a suppos'd variation of time, by which the mind can trace it in the different periods of its existence, without any break of the view . . .

(T:199–201)

The “second part” of Hume’s “system” turns on a common element as between identity (as he has defined it) and constancy, and invokes a psychological principle concerning the effects on our minds of resemblance. When an item that we may call “X” that exists at a time t_1 is identical with an item that we may call “Y” that exists at a time t_2 , then of course, given Hume’s notion of identity, the item called “X” must be exactly like the one called “Y;” for otherwise “X” and “Y” could not denote the same *unchanging* thing and so would each denote a different thing rather than a single thing. But likewise, when a set of impressions, I , had at a time t_1 bears the relation of constancy to another set of impressions, I' , had at a time t_2 , the impressions in I exactly resemble those in I' . Thus, we can say that I and I' satisfy one of the two conditions required for I and I' to be identical with each other: they are exactly alike. However, they do not satisfy the other condition required for identity, since there is an “interruption” or “gap” between them – a time between t_1 and t_2 when I and I' do not exist or occur. Were there no such interruption, we could say that I and I' were identical; in such a case, says Hume, “the interruption of our perceptions . . . is the only circumstance that is contrary to their identity” (T:209).

However, Hume also maintains that whenever A exactly resembles B, the human mind has a powerful propensity to think that A is identical with B (T:61). This is one of the psychological principles that Hume insists upon in his *Treatise*; it has for him much the same status as the principle that whenever A is regularly followed by B, we come to expect B whenever we perceive A. Hume explains the principle about resemblance by noting that observing an exact resemblance between A and B *feels like* observing an uninterrupted and unchanging object; or, as he puts it, “the act of the mind [involved] in surveying a succession of resembling objects [itself resembles] that [act of mind involved] in surveying an identical object” (T:205n). Hume thinks that this resemblance between the experience of perceiving a succession of resembling objects and the experience of perceiving a single unchanging object causes us to confuse those two different experiences and, therefore, to confuse also what they are experiences of. It is just this confusion between resembling but numerically distinct items, and a single unchanging item, that Hume ascribes to the imagination: the confusion is, according to him, the work of the imagination. But from the principle that whenever A exactly resembles B, we come to think that A is identical with B, together with the point that I and I' (sets of impressions related by “constancy”) are exactly alike, it follows that in such a case we will believe that I and I' are identical! This belief is presumably Hume’s system’s “second part;” he describes it this way:

[T]he constancy of our perceptions makes us ascribe to them a perfect numerical

identity, tho' there be very long intervals betwixt their appearance, and they have only one of the essential qualities of identity, *viz. invariableness*.

(T:202)

The upshot is that the constancy of our impressions generates a conflict or “contradiction” in our minds. On the one hand, the resemblance between impressions exhibiting constancy causes us to believe that they are identical. On the other hand, the interruptions or gaps between those impressions means that they cannot be identical. Now the “third part” of Hume’s “system” is that we resolve this conflict by “feigning” or positing the existence of something that eliminates or fills the gaps between the interrupted impressions; we come to think of those impressions not only as exactly alike (“invariable”) but also as continuous (“uninterrupted”). This allows us to think that the impressions meet both of the conditions required for identity, and are, therefore, identical. As Hume puts it:

The smooth passage of the imagination along the ideas of the resembling perceptions makes us ascribe to them a perfect identity. The interrupted manner of their appearance makes us consider them as so many resembling, but still distinct beings, which appear after certain intervals. The perplexity arising from this contradiction produces a propension to unite these broken appearances by the fiction of a continu'd existence, which is the *third* part of the hypothesis I propos'd to explain.

(T:205)

But of course, to posit or to “feign” the existence of something that eliminates the gaps between our impressions is to regard those impressions as continuing to exist while we are not perceiving them.

At this stage of his account, then, Hume has come very close to his goal of explaining the belief in body. To complete the account, however, he connects it, in the “fourth member of this system” (T:208), with a definition of *belief* that he had offered earlier in the *Treatise*: belief is “a lively idea related to or associated with a present impression” (T:96). This definition of belief is rooted in Hume’s view of causal inference: when we have experienced a constant conjunction between A and B, and have a present impression of A, our expectation of B consists in “a lively idea” of B, and Hume calls such an idea a “belief.” Hume is now saying that the belief in body also consists in a lively idea related to a present impression. This works as follows. “Our memory presents us with a vast number of instances of perceptions perfectly resembling each other, that return at different distances of time, and after considerable interruptions” (T:208).

The point here is that, since constancy consists in resemblances between sets of impressions that occur at different times, including past times, it can be recognized only by the use of memory. Putting it still differently, the material on which the imagination works when it

causes us to consider resembling sets of impressions as identical consists at least partly of present memories of past impressions (along perhaps with present sense impressions that resemble the past ones). Now Hume regards those memories as themselves consisting of impressions rather than merely of ideas, because “ideas of memory . . . are equivalent to impressions” (T:82). This “equivalence” stems, presumably, from the fact that ideas of memory, unlike purely imaginary ideas, have the liveliness in terms of which Hume defines impressions. In any case, it is these present memory-impressions that trigger our propensity to regard past – remembered – sets of resembling impressions as identical. But this propensity is so strong that, when we posit or “feign” the continued existence of those impressions, this posit has the liveliness or vivacity that characterizes belief. And since this posit is triggered by our present memory-impressions, it counts as a “lively idea related to a present impression.” Therefore, it amounts to a *belief*, namely belief in the continued existence of our impressions – which is the very belief that Hume wanted to explain. Here is Hume’s own summary of the entire account:

Our memory presents us with a vast number of instances of perceptions perfectly resembling each other, that return at different distances of time, and after considerable interruptions. This resemblance gives us a propensity to consider these interrupted perceptions as the same; and also a propensity to connect them by a continu’d existence, in order to justify this identity, and to avoid the contradiction, in which the interrupted appearance of these perceptions seems necessarily to involve us. Here then we have a propensity to feign the continu’d existence of all sensible objects; and as this propensity arises from some lively impressions of the memory, it bestows a vivacity on that fiction; or in other words, makes us believe the continu’d existence of body.

(T:208–9)

5 The “vulgar system” and the “philosophical system”

In a nutshell, Hume’s explanation of our belief in the existence of body is that the imagination, stimulated by the constancy (and coherence) of our impressions, causes us to ascribe to those very impressions a continued existence, despite the fact that they are really discontinuous, “interrupted,” or “broken.” We simply gloss over the fact that our impressions have a discontinuous existence, for the imagination causes us to confuse the impressions with continuously existing bodies. As Hume puts it: “The very image, which is present to the senses, is with us the real body; and ’tis to these interrupted images we ascribe a perfect identity” (T:205). And again:

’Tis certain, that almost all mankind, and even philosophers themselves, for the greatest part of their lives, take their perceptions to be their only objects, and

suppose, that the very being, which is intimately present to the mind, is the real body or material existence. 'Tis also certain, that this very representation or object is suppos'd to have a continu'd uninterrupted being, and neither to be annihilated by our absence, nor to be brought into existence by our presence.

(T:206–7)

Hume calls this view, which confuses discontinuous, interrupted or intermittent perceptions (specifically, impressions of sensation), with continuous, enduring or persisting bodies, “the vulgar system.” Here the term “vulgar” does not signify the modern sense of *vulgarity*; rather, it signifies that the view in question is the ordinary common-sense view – the one that humans naturally and spontaneously accept, prior to any philosophical reflection.

Hume has no sooner expounded the vulgar system, however, than he proceeds to argue that it will not withstand rational scrutiny. To show this, he simply invokes the “experiments,” described earlier, “which convince us, that our perceptions are not possess of any independent existence” (T:210). Pressing one’s eye with a finger and seeing double, or seeing a table’s perceived shape and size vary as we move nearer to or further from it, as well as “an infinite number of other experiments of the same kind,” show that “the doctrine of the independent existence of our sensible perceptions is contrary to the plainest experience.” And since, as we have seen, the continuous existence of these “sensible perceptions” (i.e. objects of sense perception) implies their independent existence – since they can have a continuous existence *only if* they have an independent existence – “this leads us back on our footsteps to perceive the error in attributing a continu’d existence to our perceptions” (T:210). But do we therefore give up our belief in the existence of body? Not at all. Instead, we introduce the philosophical theory of “the double existence of perceptions and objects.” This theory, which is none other than the causal or representational theory of perception favored by Descartes, Locke and most of Hume’s contemporaries, admits that our perceptions are “interrupted, and perishing, and different at every return.” But, at the same time, it holds that these fleeting perceptions are caused by, and represent, independently existing bodies which are “uninterrupted, and [which] preserve a continu’d existence and identity” (T:211). Hume dubs this theory “the philosophical system.”

To see the logic of the situation, consider the following three propositions:

- (1) The objects of the senses are bodies.
- (2) The objects of the senses have a continuous existence.
- (3) Bodies have a continuous existence.

The *vulgar* system, or common sense, holds that all three propositions are true. Hume, however, insists that (2) is false, and that common sense fails to recognize (2)’s falsity. He gives the complicated theory expounded in part 4 to explain why the common person wrongly affirms (2). The *philosophical* system, by contrast, recognizes that (2) is false, and

is accordingly faced with the threat of contradiction, since the negation of (2) – the proposition that the objects of the senses have a discontinuous existence – combined with (1) and (3), entails that the objects of the senses have both a continuous and a discontinuous existence. In order to avoid the contradiction, the philosophical system denies (1), by coming up with the theory of “the double existence of perceptions and objects.” The philosophical system, then, affirms the following three propositions:

- (1) The objects of the senses are our own impressions of sensation.
- not (2) The objects of the senses have a discontinuous existence.
- (3) Bodies have a continuous existence.

We already know Hume's verdict concerning the vulgar system: although it is the one that our human psychology impels us to accept, it is *false*, indeed grossly false; for proposition (2) is false, and it is only the workings of our imagination that cause us to affirm it.

What then of the philosophical system: is it any better off than the vulgar one? Hume's answer is a resounding “No;” indeed, he argues brilliantly that the philosophical system is even worse off than is the vulgar one. He announces this thesis in these words:

But however philosophical this new system may be esteem'd, I assert 'tis only a palliative remedy, and that it contains all the difficulties of the vulgar system, with some others, that are peculiar to itself. There are no principles either of the understanding or fancy, which lead us directly to embrace this opinion of the double existence of perceptions and objects, nor can we arrive at it but by passing thro' the common hypothesis of the identity and continuance of our interrupted perceptions. Were we not first perswaded, that our perceptions are our only objects, and continue to exist even when they no longer make their appearance to the senses, we shou'd never be led to think, that our perceptions and objects are different, and that our objects alone preserve a continu'd existence. “The latter hypothesis has no primary recommendation either to reason or the imagination, but acquires all its influence on the imagination from the former.”

(T:211)

Hume is here announcing his objections to the philosophical system: first, it has no rational justification; second, it derives all of its plausibility from the vulgar system. We have already seen, in effect, why he holds that the philosophical system has no rational justification: the causal inference from perceptions to objects on which that system rests is completely worthless – as part 2 of this chapter made clear. So let us now consider why he holds the philosophical system to be completely dependent on the vulgar system.

Hume's reasoning goes as follows. At first, the philosopher – the person who reflects critically on his or her beliefs, considered before the critical reflection begins – believes like

everyone else that the very objects he or she perceives *are* bodies. Upon reflection, however, the philosopher realizes that the mind-dependence of the objects of the senses, as well as the interrupted or intermittent existence that this mind-dependence entails, give the lie to this view. But, at the same time, the imagination continues to give the philosopher (in the complicated way that we outlined in part 4) an overpowering propensity to believe that the objects perceived have the continuous existence characteristic of bodies. The philosopher is thus in danger of falling into the contradiction of holding that the objects of the senses are both continuous and discontinuous. Finally, the philosopher avoids the contradiction by means of the following artifice: ascribe the interruptedness or discontinuity to perceptions, then postulate body and ascribe the continued existence to it. The contradiction is avoided by ascribing the opposed properties to different things. Thus is born the system of “the double existence of perceptions and objects,” and with it the very distinction between perceptions and objects, in the only form that Hume recognizes. But this does not alter the fact that the postulation of body has no rational foundation; it derives all of its force from the imagination’s propensity to ascribe a continuous existence to interrupted, discontinuous perceptions, except that the continued existence is now ascribed to unperceivable and unknowable bodies that are supposed to be distinct from the perceptions. The result is that the philosophical system derives all of its appeal from the vulgar system – a system which is not only false in its own right, but which actually *contradicts* the philosophical system (inasmuch as the vulgar system affirms, while the philosophical system denies, the continuous existence of our “perceptions”)! Hume’s eloquent account of this result is worth quoting:

This philosophical system, therefore, is the monstrous offspring of two principles, which are contrary to each other, which are both at once embrac’d by the mind, and which are unable mutually to destroy each other. The imagination tells us, that our resembling perceptions have a continu’d and uninterrupted existence, and are not annihilated by their absence. Reflection tells us, that even our resembling perceptions are interrupted in their existence, and different from each other. The contradiction betwixt these opinions we elude by a new fiction, which is conformable to the hypotheses both of reflection and fancy, by ascribing these contrary qualities to different existences; the *interruption* to perceptions, and the *continuance* to objects. Nature is obstinate, and will not quit the field, however strongly attack’d by reason; and at the same time reason is so clear in the point, that there is no possibility of disguising her. Not being able to reconcile these two enemies, we endeavour to set ourselves at ease as much as possible, by successively granting to each whatever it demands, and by feigning a double existence, where each may find something, that has all the conditions it desires. Were we fully convinc’d, that our resembling perceptions are continu’d, and identical, and independent, we shou’d never run into this opinion of a double existence; since we shou’d find satisfaction in our first supposition, and wou’d not look beyond. Again, were we fully convinc’d, that our perceptions are dependent, and

interrupted, and different, we shou'd be as little inclin'd to embrace the opinion of a double existence; since in that case we shou'd clearly perceive the error of our first supposition of a continu'd existence, and wou'd never regard it any farther. 'Tis therefore from the intermediate situation of the mind, that this opinion arises, and from such adherence to these contrary principles, as makes us seek some pretext to justify our receiving both; which happily at last is found in the system of double existence.

(T:215–6)

Hume's overall position with regard to the belief in the existence of body, then, is profoundly sceptical. He has examined two versions of this belief: the vulgar system on which our "perceptions" are continuous (uninterrupted) and identical with bodies; and the philosophical system on which our "perceptions" are discontinuous (interrupted) and distinct from bodies. The vulgar system, though it is supposed to be the one humans spontaneously and naturally accept, and to reflect ordinary common sense, is simply false. As Hume puts it:

'Tis a gross illusion to suppose, that our resembling perceptions are numerically the same; and 'tis this illusion, which leads us into the opinion, that these perceptions are uninterrupted, and are still existent, even when they are not present to the senses. This is the case with our popular ["vulgar"] system.

(T:217)

As for the philosophical system, not only does it lack rational foundation, but it contradicts the very system that it depends upon or "lives off.":

And as to our philosophical one, 'tis . . . over-and-above loaded with this absurdity, that it at once denies and establishes the vulgar supposition. . . . What then can we look for from this confusion of groundless and extraordinary opinions but error and falsehood? And how can we justify to ourselves any belief we repose in them?

(T:217–8)

6 A criticism of Hume's account

The thesis that leads Hume to contend that common sense is mistaken in holding that the objects of the senses have a continuous existence is, as we have seen, the philosophical view that the objects of our perceptions are our own impressions. As I have said, this thesis is the most basic of the three assumptions that govern Hume's whole account of the belief in the

existence of body, and it was widely accepted by his contemporaries. Thus far, the thesis has not been examined critically; we have confined ourselves to tracking its role in Hume's account. The time has come, then, for us to examine this key thesis critically. Now there is a huge philosophical literature on this topic; indeed, perhaps no other question has been debated as intensively in twentieth-century philosophy of perception as this one: *do we perceive material things, or do we perceive only impressions* (or, as these have been variously called, sensations, ideas, percepts, appearances, *sensa*, sense-data)? Until about the middle of this century, most philosophers believed that arguments such as those sketched at the beginning of this chapter force us to accept the latter view, which is now commonly called the "sense-datum theory." Since that time, however, the arguments for the sense-datum theory have been intensively criticized, and most (though not all) philosophers now reject both the arguments and the theory. We cannot canvass here all of the arguments for the sense-datum theory.⁷ Rather, we shall return only to those arguments, surveyed in part 2 of this chapter, that Hume advances when he speaks of the "experiments, which convince us, that our perceptions are not possess of any independent existence:" namely, the argument from perceptual relativity and the argument from the illusion of seeing double; and, more briefly, to the epistemological argument that appeals to the many alternative ways of causing any perceptual experience. We shall argue that those arguments are faulty, and our objections to them will be, in content and spirit, characteristic of the objections that have led most contemporary philosophers to reject the sense-datum theory.

Here then, once again, is Hume's brief statement of the argument from *perceptual relativity*:

The table, which we see, seems to diminish, as we remove farther from it; but the real table, which exists independent of us, suffers no alteration: it was, therefore, nothing but its image, which was present to the mind.

(E:152; S:104; F:183)

One way to summarize this reasoning is as follows:

- (1) When we look at an object from different distances and angles, what we see changes.
 - (2) When we look at an object from different distances and angles, the object itself does not change.
-
- ∴ (3) When we look at an object from different distances and angles, what we see is something other than the object itself – an impression, image, or sensedatum.

Stated in this fashion, the argument is logically valid, but are both its premisses true? Many philosophers today, including myself, would say that while premiss (2) is obviously true, (1) is simply false. What we see on looking at an object from different distances and angles does not *really* change; rather, it only *seems* to change. In other words, (1) ought to be replaced with:

- (1a) When we look at an object from different distances and angles, the object's size and shape *seem* to change.

Unlike (1), (1a) does not assert that there is a seen object that really changes when we look at an ordinary object under different conditions; it says only that the object itself seems (in one sense of "seems") to change – that its size and shape *look* different from different distances and angles.⁸ Hume's own language suggests this way of putting it: he says that what we see "*seems* to diminish" as we move away from it, not that it *does* diminish. The introduction here of the little word "seems" makes all the difference. For, once (1a) is substituted for (1) in the argument, it is no longer valid: from (1a) and (2), all that follows is that the object itself *looks* or *seems* different from different distances and angles, which is exactly what we expect (what would be strange is if the object continued to look exactly the same from any distance and angle!), and from which it does not follow at all that the object we see from one distance and angle really is different from the object we see from another distance and angle. The upshot is that the argument is either valid but unsound – the case where premiss (1) is retained – or invalid – the case where (1a) is substituted for (1).⁹

Some contemporary philosophers have given an illuminating diagnosis of the basic error committed by the argument from perceptual relativity. The argument's proponents have assumed that if a premiss like (1a) is true, then a premiss like (1) must also be true. More basically, they have assumed that from a premiss of the form: (A) Person S perceives something that *seems* or *appears* *F*, there follows a conclusion of the form: (B) Person S perceives an appearance (impression, sensedatum) that really *is* *F*. Suppose for example that someone, say Mary, sees a coin that, to her, looks elliptical because of her angle of vision. Then there does seem to be an elliptical *something* in her visual field, and so philosophers have assumed that Mary sees an object which really is elliptical. But since the coin is round, the elliptical object that Mary supposedly sees cannot be the coin itself. What then is it? The answer, it seems, is that this object can only be an impression – a percept, "image," or sense-datum. If this mode of reasoning were valid, then from a premiss like (1a), which says that what we see under different conditions *seems* or *appears* different, one could always derive a premiss like (1), which says that what we see under different conditions really *is* different.

In fact, however, the assumption that a premiss of form (A) entails a conclusion of form (B) is wrong. This point has been especially well made by Roderick M. Chisholm, who is an influential contemporary epistemologist and philosopher of perception. Chisholm shows that there are many arguments with a form-(A) premiss and a form-(B) conclusion that are obviously invalid – that can obviously have a true premiss and a false conclusion. Two such arguments are:

John sees a dog which looks vicious and more than 10 years old.

∴ John sees an appearance (impression, sense-datum) which is vicious and over 10 years old.

Mary sees a man who appears tubercular.

∴ Mary sees an appearance (impression, sense-datum) which is tubercular.¹⁰

Chisholm calls any argument with a form-(A) premiss and a form-(B) conclusion a case of “the Sense-Datum Fallacy,” on the grounds that all such arguments are invalid, and because there is no true premiss that can be added to make them valid.¹¹

Notice that our objections to the argument from perceptual relativity can also be applied to the argument from the causal facts of perception. That argument could be summarized this way:

- (1') When an object is perceived by means of different sense-receptors (e.g. the eyes of a human and those of a housefly, or the eyes of a near-sighted person and those of a person with 20-20 vision), what is perceived varies.
 - (2') When an object is perceived by means of different sense-receptors, the object itself does not vary.
-

∴ (3') When an object is perceived by means of different sense-receptors, what is perceived is something other than the object itself – it is an impression, image, or sense-datum.

In parallel with our criticisms of the argument from perceptual relativity, we can here say that premiss (1') is false. What is true is rather: (1a') When an object is perceived by means of different sense-receptors, what is perceived seems to vary. But from (1a') and (2'), (3') does not follow. So again, the argument either has a false premiss, or is invalid.

Let us turn next to Hume's argument from *the illusion of seeing double*; here, once more, is his presentation of it:

When we press one eye with a finger, we immediately perceive all the objects to become double, and one half of them to be remov'd from their common and natural position. But as we do not attribute a continu'd existence to both these perceptions, and as they are both of the same nature, we clearly perceive, that all our perceptions are dependent on our organs, and the disposition of our nerves and animal spirits.
(T:210–11)

We may reconstruct this argument somewhat informally, as follows:

- (1) When we press one eye with a finger, we see two of every object that we previously saw.
- (2) At least one member of every such pair of objects lacks a continuous and independent existence.

(3) Both members of every such pair of objects are of the same nature.

∴ (4) Both members of every such pair of objects lack a continuous and independent existence.

Although this argument may have a certain plausibility, it is vulnerable to at least two objections. One objection concerns premiss (1), which says that upon pressing one eye we really see *two* objects for every one that we saw before – that the number of objects now seen actually increases (it doubles). The objection is that this is to take the element of truth expressed in premiss (1) too literally. All that premiss (1) really commits us to is that upon pressing one eye, we *seem* to see a “twin” of every object that we previously saw. But *seeming to see* a new object is not the same thing as *actually seeing* a new object – as would happen if, say, the object had undergone mitosis or fission – and the argument simply confuses these two importantly different things.

The second objection is that even if we were to grant premiss (1), and so to grant also the talk in (2) and (3) of pairs of “objects” that are seen when pressing one eye with a finger, the argument would still be faulty; for there is no good reason to accept (3), the claim that both members of such a pair of objects are “of the same nature.” For why could we not maintain, instead, that one of the two objects in such a pair – the one that is not “remov’d from [its] common and natural position” – is a body or material thing, while only the other is an impression or “perception”?

Hume himself gives no reason why we should not say this, but more recent proponents of the sense-datum theory would give one on his behalf. This is that there is no special, discernible, tell-tale difference between the two “objects” to indicate that the one is a body, while the other is a fleeting, mind-dependent, impression; therefore, they must both be of the same nature, just as (3) says.¹² Thus, since according to (2) at least one of them must be a fleeting and mind-dependent impression, both must be impressions, just as the conclusion states.

But this reason for saying that both “objects” are of the same nature is defective. For the fact that two things are visually alike – that we can see no discernible difference between them – does not at all prove that they are “of the same nature.” A real tomato and a wax replica of a tomato may be visually indistinguishable, but that certainly does not show that they are the same in nature. As J. L. Austin put the point in his short but very influential book *Sense and Sensibilia* (1962: 50–2):

If I am told that a lemon is generically different from a piece of soap, do I “expect” that no piece of soap could look just like a lemon? Why should I? . . . Why on earth should it *not* be the case that, in some few instances, perceiving one sort of thing is exactly like perceiving another?

We may conclude that Hume’s argument from the illusion of seeing double is unsound: its

first premiss is false and, even if that premiss were accepted, there would be no reason to accept its third premiss.

In part 2 of this chapter it was said that certain *epistemological arguments* have motivated the view that we perceive impressions or sense-data rather than physical objects. One such argument, we saw, stems from the consideration that any perceptual experience obtained when a physical object is stimulating one's sense-receptors could be caused in other ways, including direct brain stimulation or even an "evil deceiver" like the one postulated by Descartes in his *First Meditation*. We said that because this possibility threatens our deep-seated conviction that perception is a source of knowledge, some philosophers have suggested that in perception we come to know at least the existence and nature of our own impressions (from which we then have to infer the existence of some physical object). While this argument (to which we shall return in the next part) is a powerful one, the point needs to be made here that the argument does not establish that we perceive (or "immediately perceive," as is usually said) impressions or sense-data. For it is possible to hold that in perception *something* other than the existence and nature of physical objects can be known *without* holding that this something must therefore be the existence and nature of some *other* kind of object. Rather, what is known in perception can be simply a fact about one's self – a fact that contemporary philosophers such as Chisholm propose to describe by saying that the self is "appeared to" in a certain fashion, or that the self "senses" in a certain fashion. This is a complex matter that we shall not pursue in more detail here; but the essential point is that the epistemological purpose of introducing "impressions" or "sense-data" into an account of perception – which is to allow that something can be known in every case of perception despite the fact that any perceptual experience can be caused in a variety of ways – can be secured without postulating such special objects of perception, by holding that in perception we gain knowledge of certain sensory states of the self (from which, it is argued, we can infer the existence of physical objects causing those states).¹³

7 A contemporary perspective on Hume's account

In light of the above criticisms of Hume's account, you might now conclude that there is little to be learned from it, other than the avoidance of certain errors, such as the "sense-datum fallacy." But, to draw this conclusion would be too hasty; for it would be to assume that the sceptic's problems with sense perception depend wholly on the doctrine that we perceive only impressions, or percepts, or sense-data. In fact, however, such an assumption would be wrong; for "scepticism with regard to the senses" can arise quite independently of that doctrine.¹⁴ A concise statement of how such scepticism can be generated is given by Robert M. Adams:¹⁵

[O]ne of the most important sceptical arguments in modern philosophy . . . is due to Descartes and is based on the idea of "a God who is able to do anything."

“How do I know that he did not bring it about that there be no earth at all, no heavens, no extended thing, no figure, no size, no place, and yet all these things should seem to me to exist precisely as they appear to do now?” Descartes was no sceptic, and thought he could solve the problem by proving the existence and nondeceitfulness of God. Descartes’s problem has proved more durable than his solution, however . . .

The problem remains. All our beliefs about the material world are based ultimately on sense experience. Yet it seems quite conceivable that everything might seem to us exactly as it actually seems to us in sense perception even if there were no material world at all outside our minds. So how do we know there is a material world outside our minds?

The sceptical argument that Adams here rightly attributes to Descartes does not involve the view that we cannot perceive material things but only impressions or sense-data. Rather, the argument is designed to show that even if we can, or in fact do, perceive material things, we can never *know* that we do, because such knowledge must be based on our sense experiences (visual experiences, tactile experiences, auditory experiences, etc.); but it is perfectly conceivable that we could have the very same sense experiences that we do in fact have even if no material things existed: this could happen if God, or perhaps some other being, such as Descartes’ “evil deceiver,” were causing us to have all the experiences.

Hume himself would presumably have accepted this argument’s sceptical conclusion, since he explicitly holds that the belief in body has no rational basis. He contents himself, as we have seen, with pointing to certain features of our sense experience – constancy and coherence – that are supposed to explain causally why we believe in body. Some contemporary philosophers, however, have tried to use Hume’s ideas about how constancy and coherence give rise to the belief in body in a way that was not envisioned by Hume. These philosophers have tried to see constancy and coherence as features of sense experience that *justify* the belief in body, or, as Jonathan Bennett (1971: 313–53) puts it, that justify our use of “objectivity-concepts” (See also Ayer 1940: 243–63). It is presumably because Hume’s discussion lends itself to this purpose that Bennett, in the passage we quoted at the beginning of this chapter, says that despite being “full of mistakes and – taken as a whole – a complete failure,” Hume’s treatment is “one of the most instructive arguments in modern philosophy” (1971: 311). Let us therefore consider in rather more detail this manner of reading Hume, using Bennett as our guide.

Bennett focuses on a passage in which Hume discusses the role of *coherence*. Hume writes:

I am here seated in my chamber with my face to the fire; and all the objects, that strike my senses, are contained a few yards around me . . . I hear on a sudden a noise as of a door turning upon its hinges; and after a little see a porter, who advances towards me. This gives rise to many new reflexions and reasonings.

First, I have never observ'd, that this noise cou'd proceed from anything but the motion of a door; and therefore conclude, that the present phaenomenon is a contradiction to all past experience, unless the door, which I remember on t'other side of the chamber, is still in being But this is not all. I receive a letter, which upon opening it I perceive by the hand-writing and subscription to have come from a friend, who says he is two hundred leagues distant. 'Tis evident I can never account for this phaenomenon, conformable to my experience in other instances, without spreading out in my mind the whole sea and continent between us, and supposing the effects and continu'd existence of posts and ferries, according to my memory and observation. To consider these phaenomena of the porter and letter in a certain light, they are contradictions to common experience, and may be regarded as objections to those maxims, which we form concerning the connexions of causes and effects. I am accustom'd to hear such a sound, and see such an object in motion at the same time. I have not receiv'd in this particular instance both these perceptions. These observations are contrary, unless I suppose that the door still remains, and that it was open'd without my perceiving it: And this supposition, which was at first entirely arbitrary and hypothetical, acquires a force and evidence by its being the only one, upon which I can reconcile these contradictions. There is scarce a moment in my life, wherein there is not a similar instance presented to me, and I have not occasion to suppose the continu'd existence of objects, in order to connect their past and present appearances, and give them such an union with each other, as I have found by experience to be suitable to their particular natures and circumstances. Here then I am naturally led to regard the world, as something real and durable, and as preserving its existence, even when it is no longer present to my perception.

(T:196–7)

Bennett sees Hume to be making two different points in this passage:

- (1) Unless I admit that the door, sea, continent, post offices, and ferries, etc., continue to exist while I do not perceive them, my experiences as of now hearing a door, seeing a letter, etc., are “contradictions” of my past experience.
- (2) Unless I admit that the door, sea, continent, post offices, and ferries, etc., continue to exist while I do not perceive them, I cannot “account” for my experiences as of now hearing a door, seeing a letter, etc., in a way that conforms to my past experience.

We can illustrate the first point this way: suppose that upon having the experience of seeing the letter, I deny that the sea, continent, post offices and ferries continued to exist while I was not perceiving them. Then Bennett reads Hume as saying that I must believe both that “I received a letter that was ferried last week by the postal services across the sea and

continent” and that “the sea, continent, postal services and ferries did not exist during the past week” – which is self-contradictory. Bennett, however, rejects this point, on the ground that the contradiction can be generated only if I *already* accept the existence of such things as the sea, continent, post offices, and ferries: if all I know is that I am now having certain visual experiences as of receiving a letter, and that in the past I have had experiences as of seeing the sea, continent, posts and ferries, then there is no contradiction in denying that any of these things continued to exist while I did not perceive them. As he puts it:

The notion of “contradiction” has no place here unless I already accept a large body of theory: the proposition that I inhabit a world of objects, many hypotheses about their general behaviour, and some hypotheses of the form “I have perceptions of kind K only when in the presence of objects of kind K*”.

(Bennett 1971: 324)

This objection seems right. On the other hand, Bennett sees much more merit in Hume's second point: He writes (1971: 325):

[W]e could take [Hume's talk of removing contradictions] to mean “explain, without contradicting my other experience”. This would put Hume on firmer ground: spreading out the sea and the continent is no longer removing a contradiction but providing an explanation.

And, a little later (p. 330), Bennett even talks of [Hume's] “manifest success, in the ‘door’ passage, in showing that The Belief [in the existence of body, of material things] does have a certain kind of legitimacy.”

Bennett's position here, in its general character, is a fairly standard one in the philosophy of perception. The basic idea is that our sense experiences present certain features – especially certain kinds of order and recurrence, such as Hume's constancy and coherence – that are best explained by the theory that the experiences are experiences of an objective world, in which bodies exist independently of those experiences and causally interact among themselves and with our perceptual organs in regular, law-like ways. Bennett puts this basic idea as follows:

I have a conceptual framework which lets me connect my various sensory episodes to form a coherent whole: I bring the brute, disconnected facts of my sensory history under a *theory* in terms of which I can adduce some of these facts as explaining others, can predict further ones, and so on. This theory does its work only because, through it, statements about past perceptions can imply statements about present and future ones. Also, the theory is . . . so structured that through it

certain perception-statements can imply the existence of objects when I do not perceive them.

(1971: 325)

The manner in which Bennett spells out the details of this general position, however, is original and of considerable interest. Building on Hume's example of the door, Bennett illustrates the way in which the "theory" of independently existing objects enables me to "connect my various sensory episodes to form a coherent whole" – to adduce some of "the brute, disconnected facts of my sensory history . . . as explaining others, . . . [and as] predict[ing] further ones." He writes (p. 331):

When I see a door, turn my head away, and then turn back and see it again, my two visual impressions are connected by the statement that each is the seeing of a door. For example, the judgement that the first is the seeing of a door supports the prediction that when I turn my head back I shall have the second visual impression.

Here the point is that by treating my first visual experience as the perception of a door – a real door that exists independently of my experience of it – I can use the experience to predict that when I turn my head back I shall again have a visual experience as of a door, and also to explain why I have this latter experience upon turning my head. If, on the other hand, I regarded the first experience as just a sensory episode in my psychic history, not connected to or caused by a door, then I could not use it in these ways. Thus, Bennett declares: "In ways like this I bring objectivity-concepts to bear upon impressions which I do have; and Hume should admit that such procedures are useful, and are broadly causal in nature" (p. 331).

Having in this way illustrated how the belief in *independently* existing objects explains how we can use some experiences to explain and predict other experiences, Bennett, in one of the most interesting moves in his discussion, goes on to argue that such a belief perforce commits one also to the belief in *continuously* existing objects. He first invites us (p. 332) to suppose that Hume would resist this transition: that he would try to admit the explanatory usefulness of independently existing objects but refuse to extend this admission to continuously existing ones:

Hume might still say: "But the fact remains that the utility of 'perceived object' does not help me with my problem about the utility of 'unperceived object'. Even if it is helpful to be able to say 'I now see a door' both at t_1 and t_3 , I have shown that it cannot be helpful to say 'There is now a door which I don't see' at t_2 ."

To this, Bennett responds:

But the benefits of “perceived object” come through a theory of which “unperceived object” is an *integral* part: we cannot lop off statements asserting the existence of objects while unperceived, while retaining the “useful” objectivity-statements which classify some of our perceptions as perceptions of objects.

(1971: 332)

This seems right; for suppose that, in the door example, I did *not* admit that the door continued to exist while I was looking away. Then my first visual experience would provide no basis for predicting or explaining the second; instead, I might as well suppose that a door like the first one *miraculously* sprang into existence when I looked again, which would be as good as admitting that the belief in an independently existing door had *no* explanatory value after all. Bennett goes on (pp. 332–3) to illustrate this vital link between admitting independent objects and continuous ones in terms of Hume’s “sea and continent” example:

[T]he “sea and continent” are supposed to raise a problem for me only in respect of their existence when I don’t perceive them; but the statement “There is a sea between us *now*” is connected, through my general world-theory and thus through the conceptual framework which is its bone, with statements about perceptions which I do have at some time or other – what I shall observe if I take a certain journey, what I heard my friend say when asked “What route will you take?” and so on. I have no way of linking these perceptions of mine which doesn’t involve my agreeing that the sea is there right now I rest my case on the . . . thesis that the only way I can helpfully bring objectivity-concepts to bear on *my* impressions – or you on yours, or Hume on his – is through a theoretic structure which, together with the given experiential data, implies that objects sometimes exist while unperceived. The only ground *we* have for claiming sometimes to perceive objects are equally grounds for claiming that objects sometimes exist when we do not perceive them.¹⁶

We shall conclude this part by raising a question about Bennett’s “explanation-providing” defense of the belief in body. Clearly enough, the defense works only if a good case has been made for the belief in the *independent* existence of the objects of perception; for only then can Bennett’s chief point – that any grounds there are for saying that we perceive (independent) objects are grounds also for saying that those objects continue to exist unperceived – come into play. The question we shall raise is simply this: has Bennett made a good case for the belief in object-independence? Such a case, it seems, would have to address effectively the sceptical argument, described at the beginning of this part, which R. M. Adams traces to Descartes: it would have to give a good reason for saying that our perceptual experiences are produced not by some extremely powerful and intelligent agent bent on deceiving us but rather by a physical world. But the only reason Bennett provides is the one he describes

rather blandly, in connection with the door example, by saying: “In ways like this I bring objectivity-concepts to bear upon impressions which I do have; and Hume should admit that such procedures are useful, and broadly causal in nature” (1971: 331).

But suppose that a determined sceptic were to respond to Bennett by saying:

“I admit that one way in which I can use a present perception as of a door to predict another such perception, or use one perception of a door to explain another one, is by agreeing that a door exists independently of those perceptions and continues to exist in between the times of their occurrence. But this is only one possible explanation; there are others. Another possible explanation is that God, or some evil deceiver, wants to make me believe falsely that doors and other material things exist, and so causes in me a set of orderly and systematic perceptual experiences such that I can predict those I will have on the basis of those I have had, and explain some perceptions by adducing other perceptions, and so forth.”

Now there seems to be nothing in Bennett’s account – at least nothing that he explicitly says – to explain, even to suggest, why such an alternative theory is less reasonable than the physical-world theory.

Furthermore, in at least one passage, Bennett seems to overlook the type of sceptical challenge that we have raised. Bennett writes (p. 325) that, in the light of his account:

In a given sensory situation I may have to choose between (*a*) accepting that there is such an object [i.e. an object that exists when I do not perceive it], (*b*) relinquishing the theory and thus my only chance of explaining my sensory present, and (*c*) retaining the theory while denying that there is an object which I don’t perceive – thus committing myself to a falsehood about my sensory past. This is the case where I must choose (*a*) if I am to “account for [my present perception] 17 conformable to past experience”: if I choose (*b*) I cannot “account for” the perception, and if I choose (*c*) my explanation will not be “conformable” to my past experience.

To illustrate: suppose that my “given sensory situation” is that I seem to see a door on turning my head back, or have an experience as of seeing a door on seeming to turn my head back, along with the memory of having had a similar experience (minus the head-turning experience) a moment ago. Then, according to Bennett, I have three choices: (*a*) I accept that I am seeing a door; (*b*) I give up the theory that doors and other objects exist independently of my perceptions of them; and (*c*) I retain this theory but deny that I am now seeing a door. Bennett would say that if I choose (*b*) I relinquish “my only chance of explaining my sensory present;” and that if I choose (*c*) my explanation of my present sensory situation, whatever it is, will not be the same as it was in other similar situations; which is why, if I wish to explain my present sensory situation, and to do so in a way that matches similar

situations (and, otherwise, what sort of “explanation” would it be?), I must opt for (a). The trouble, however, lies with Bennett’s claim that if I opt for (b) then I relinquish “my *only* chance of explaining my sensory present,” and so “cannot ‘account for’ the perception [emphases added].” A determined sceptic will not grant this claim. Rather, such a sceptical philosopher will say that there are other possible theories, such as the theory of a universal deceiver, that can serve to explain my present sensory situation just as well as the physical-world theory. There seems to be nothing in Bennett’s account that can effectively meet this classic sceptical challenge.¹⁸

8 A Kantian response to scepticism regarding the senses

In this final part, we shall consider briefly another, quite different, response to the sceptical challenge: that of Immanuel Kant in his *Critique of Pure Reason* (hereafter, “the *Critique*”). As previously mentioned, Kant’s theory of knowledge is an enormously complex one that lends itself to many different interpretations; here we shall consider only one key idea in *one* way of reading Kant.

As was mentioned in Chapter 5, in the *Critique* Kant gives a number of “transcendental arguments” – arguments intended to show that we could not have the kind of experience we do, unless certain principles were true. Such arguments start from some fact about experience which is supposed to be incontrovertible and which even a sceptic would admit, and then try to show that this fact, when reflected upon, implies the truth of a principle(s) that the sceptic doubts. The most ambitious such argument offered by Kant, in a famous and very difficult section of the *Critique* called the “Transcendental Deduction of the Pure Concepts of the Understanding,” can be interpreted as an attempt to show that a certain incontrovertible fact about human experience would not be possible unless we employed concepts of physical objects.

On the interpretation of Kant’s transcendental deduction (as this argument is usually called) that we shall briefly consider, the incontrovertible fact of experience with which the argument begins is what we shall call the “unity of consciousness.”¹⁹ To see what this means, notice first that human consciousness at any given time is typically not consciousness of just one thing, but rather of *many* things. We are seldom if ever conscious of only one item – our experience would be very dull if we were – but of a diversity of items (which Kant called a “manifold of representations”). For example, when simply gazing into an ordinary furnished room, you are conscious not only of a door, but also of walls, a chair, a table, a bed, windows, a carpet, light fixtures, and so on. Now this simple fact requires something noteworthy; namely, that each of the items of which you are aware be presented to *one and the same* consciousness. If the chair were presented to one consciousness, the table to another, the bed to a third, and so on, then there would be no awareness of the furnished room, but only separate awarenesses of a wall, table, chair, bed, etc. This point, as we saw in Chapter 1 – in connection with our (Kantian) objection to Hume’s bundle theory of the self – is also evident

when we recall that human consciousness is extended over time, and that we can be aware of succession in time. In order to hear several strokes of a bell as a succession of strokes, it will be recalled, it is essential that each stroke be retained in one and the same consciousness when its successor is heard.

Kant's transcendental deduction – or at least one key idea in one interpretation of his highly complex argument – turns on the relation between this fact of the unity of consciousness on the one hand, and the use of concepts of objects on the other. Notice first that unless we had unity of consciousness, we could not use concepts of objects. Consider a simple object like a chair. If each leg of the chair, the seat of the chair, and the back of the chair were always presented to different consciousnesses, then there could be no consciousness of a chair, and so no concept of a chair. Concisely put, consciousness of an object requires unity of consciousness; the parts of the object of which we must be aware in order to be aware of the object, be they few or many, must each be given to one and the same consciousness. Now the key idea in Kant's argument is both less obvious and more difficult; it is the converse of the claim just made: to wit, that unity of consciousness requires consciousness of an object.

To see why Kant thinks this is so, suppose that we ask: what unifies, or holds together in a single consciousness, a variety of experiences? This is not an easy question to answer, for there are at least four possible answers that must be eliminated:

- 1 It will not do to answer that a variety of experiences can be unified by introspectively spotting some relation that they all have to the self or conscious subject. This is because, as we saw in Chapter 1, Hume showed that we do not and cannot introspect the conscious subject – a point also made by Kant (1963: A107). It follows that we cannot introspect the self *in relation* to items of experience, since to be introspectively aware of a relation between X and Y we must be aware at least of both X and Y.
- 2 A variety of experiences cannot be unified by spotting some necessary connections between the experiences themselves. For, as we saw in Chapter 4, Hume showed that we never observe any necessary connections between different occurrences.
- 3 A variety of experiences cannot be unified by relations of association between them.²⁰ This is because, as the Kant scholar Robert Paul Wolff has shown (1963: 108–9), mere association is not sufficient for unity of consciousness. For, suppose that John and Mary's experiences have the following unusual relationship: whenever John smells bacon, Mary imagines eggs; and whenever Mary imagines eggs, John smells bacon. Then Mary's and John's experiences of smelling bacon and imagining eggs stand in a relation of association – one leads to the other and vice-versa – but they do not belong to one and the same consciousness. For it is not the case that when John smells bacon, *he* imagines eggs; or that when Mary imagines eggs, *she* smells bacon. To put it differently, the bacon-smelling and egg-imagining experiences are not in the same mind.
- 4 We cannot say that a diversity of experiences are unified in one consciousness because they are all in the head (see Wolff 1963: 115). For, even waiving objections to talking about heads in the context of trying to meet the sceptical challenge, as well as possible

objections to talk of experiences being “in the head,” this condition would not be sufficient: the experiences might be in different parts of the brain that did not communicate with each other.

If a diversity of experiences cannot be unified in any of the above four ways – by standing in an introspectional relation to the self, by exhibiting necessary connections between each other, by association, or by being all in the head – then what does unify them? The answer proposed by Kant is this: the experiences are unified by being related to an object, by being *of* an object. To grasp this idea better, notice that being related to an object is certainly one way to unify a diversity of experiences: the object serves, so to speak, as an anchor for them. Now Kant's idea is that, in light of the elimination of the other possibilities mentioned, this is the *only* way to unify a diversity of experiences in one consciousness (Wolff 1963: 116).

This idea is merely the nub of a lengthy argument that we shall not develop more fully here, since our purpose is merely to sketch the core of one reading of Kant's alternative response to an external-world sceptic.²¹ Rather, let us test Kant's idea, by considering some possible objections to it.

One possible objection would be that Kant is wrong to think that relation to an object is required to unify a diversity of experience, because mere *co-presence* to one consciousness is sufficient to explain such unity. Kant could reply that this objection just begs the question; for “co-presence” is here merely another term for “unity of consciousness” – the very notion he is trying to explain. Further, co-presence may seem to provide the needed explanation, only because we can think schematically of a set of items that are all related to the same self by (the relation of) awareness. However, once we realize that we cannot spot any such self in introspection, we can see that co-presence offers no explanation of the difference between a set of experiences each of which exists for a different consciousness, and a set of experiences each of which exists for one and the same consciousness. Rather, it is only a label for the latter phenomenon, which is the one we are trying to understand.

A second possible objection goes as follows. What would Kant say about the experiences obtained in a hallucination? Does not his idea “prove too much,” namely, that even hallucinatory experiences are experiences of objects? Kant could reply that even hallucinatory experiences *purport* to be of objects, they are conceptualized *as* experiences of objects. It is only because they do not cohere with other experiences, or with the broader context of our experience, that they are eventually classified as hallucinations.

This reply, however, leads to another, more basic, possible objection to Kant's approach: does not his idea show only that our experiences must be *conceptualized as* experiences of objects, while failing to show that they must really *be* experiences of objects? If so, then it would seem that Kant's approach completely fails to meet the sceptical challenge.

To see how Kant could answer this basic objection, we need to describe the question to which the constructive part of the *Critique* is addressed. This question can be put as follows: what can we know about what human experience (which is the only kind we can have) must be like, without prejudging or even addressing the question of whether that experience

conforms to things as they are quite apart from our experience – indeed without even addressing the question of whether things as they are in themselves are the same things as are different from the things we experience? For Kant, this question breaks down into two:

- (1) What can we know about how anything must be *perceived* by us, without making any judgment on the question of whether this is the same or different from the way things are apart from the way we perceive them?
- (2) What can we know about how we must *think* things are, or conceptualize them, without making any judgment on the question of whether the way we must conceptualize things conforms to things as they may be apart from such conceptualization?

To the first question, Kant answers, in the first constructive part of the *Critique*, called the “Transcendental Aesthetic,” that things must be perceived by us as being in time and space. To the second question, he answers, in the second constructive part of the *Critique*, called the “Transcendental Analytic,” that given just the fact of unity of consciousness as an incontrovertible datum, it follows that things must be conceptualized as physical objects (whose changes are, moreover, governed by causal laws).

We cannot explore in this work the deep and complex arguments by which Kant seeks to defend these answers to his questions. But let us try to say how his approach bears on the sceptical challenge. Notice that Kant’s questions open up a kind of space in which we can talk about human experience without the worry of whether it conforms to things as they may be in themselves – things apart from our modes of experiencing them. Kant would say that this is in fact the proper domain of philosophy, and that both the non-sceptic and the sceptic make the same mistake. The non-sceptic claims that we can have knowledge of things as they are in themselves, quite apart from our ways of experiencing them. This, by Kant’s lights, is a mistake; for we cannot possibly know what things are like apart from the ways in which we must perceive and think of them. As for the sceptic: it might, at first, be thought that Kant is himself a sceptic; for does not a sceptic hold precisely what Kant holds – that we cannot know what things are like in themselves? But there is a difference between Kant and the sceptic. A sceptic holds that we cannot know what things are like in themselves because our *evidence* or *justification* for any claim about such things is always insufficient or faulty, as in the case of the causal inference from impressions to objects, which, as we saw, Hume rejects. Kant, by contrast, regards as incoherent the very thought that there could be evidence as to what things are like completely apart from the ways in which we must perceive and conceptualize them. For that thought fails to recognize that in all human cognition, there is a contribution made by the knower: namely, the knower’s ways of perceiving and conceptualizing things, the import of which for the content of our knowledge cannot be ignored or factored out.

The mistake made by sceptic and non-sceptic alike, then, is to assume that it makes sense to suppose that we could know what things are like “in themselves” – independently of the

ways in which we must perceive and conceptualize them. In particular, the sceptic's supposition that all our experience might be produced by a Cartesian deceiver is precisely a supposition about what reality might be like totally apart from the ways we must perceive and think of it. So Kant presumably would have to grant that we cannot know this supposition to be either true or false, for that would again be to know what things are like totally apart from our ways of perceiving and conceptualizing them. But he would maintain that we can at least know that we must perceive and conceptualize reality as a system of physical objects interacting causally in space and time. Whether this is a satisfactory reply to the sceptical challenge must here be left for the reader to ponder.²²

NOTES

INTRODUCTION

- 1 As Hume scholars have shown, there were still other formative influences on Hume. Norman Kemp Smith especially, in his *The Philosophy of David Hume* (1941), has emphasized the influence of the British moralists Hutcheson, Shaftesbury, Mandeville, and Butler, especially Hutcheson.
- 2 This general account of Hume's basic program is indebted in part to D. G. C. MacNabb 1967: 75.

1 HUME'S THEORY OF MEANING AND ITS IMPLICATIONS

- 1 This summary of Descartes' presentation of the case for scepticism follows his *First Meditation*. See Descartes 1984: 12–15.
- 2 This topic will be discussed in more depth in Chapter 6.
- 3 Despite the advantages of the modern formulations, some contemporary scholars would resist the attempt to interpret Hume along these lines. For example, Garrett (1997: Chapter 2) argues that:
 - (1) Hume did not regard the “copy principle” (as Garrett calls Hume's version of the principle of empiricism) as an a priori truth (p. 43); and
 - (2) some of Hume's principles concerning the operations of the imagination offer him a way of explaining how one might have the idea of the missing shade of blue without having had the corresponding impression (pp. 51–2).

With regard to (1), it should be noted that Garrett's principal argument – that (a) impressions are *causes* of ideas and (b) for Hume causal claims are never a priori – is based exclusively on the *Treatise*. In the *Enquiry*, Hume does not insist or even say that impressions are causes of ideas. Given that Hume himself says in his prefatory “advertisement” to the *Enquiry* that it alone should be regarded as containing his “philosophical principles and sentiments,” that he there also distances himself from the *Treatise* by calling it “that juvenile work” (E:2; S:xviii, F:52), and that one of the main differences between the two works is precisely that the *Enquiry* contains far less psychology, this difference between the accounts of impressions and ideas in the two works is not insignificant. With regard to (2), it should be noted that (as Garrett himself seems quite aware) explaining how one might, on Humean principles, arrive at the idea of the missing shade of blue is not tantamount to showing that there is no contradiction between the copy principle and Hume's admission that this idea need not be copied from any impression. Despite these reservations about Garrett's position, however, we need not deny that he may be right in claiming that his reading of Hume is historically accurate. For our claim is not that the modern formulations are *identical* with Hume's own views. Rather, it is that:

- (1) the modern formulations are suggested by, and historically descended from, Hume's own views;
 - (2) they do not face the difficulties that afflict Hume's own views; and
 - (3) they can do the same critical work regarding "substance," "self," and "cause" that Hume's own principle of empiricism and test for meaning are supposed to do.
- 4 Since philosophers routinely distinguish between *properties* (e.g. redness, roundness), which are abstract or general and cannot be said to exist at one place and time rather than another, and *property-instances* (e.g. the redness and roundness of a particular apple) which are particular and can be said to exist or occur at a particular place and time, it would be more accurate to say that, according to the bundle theory, a thing is composed of coexisting property-instances (or coexisting instantiated properties, or co-instantiated properties), while according to the substance theory, it is composed of property-instances that belong to an underlying substance. But, for simplicity's sake, we shall continue to say that on the bundle theory a thing is composed of coexisting properties.
 - 5 Descartes held that the basic or "primary" properties or qualities of matter are restricted to shape and size, which he calls "extension"; Locke argued that solidity must also be included, so as to distinguish matter from empty space. Both Descartes and Locke held that the so-called "secondary" qualities or properties – color, taste, smell, sound, and temperature – are merely capacities or dispositions that objects have, because of their primary properties/qualities, to produce certain experiences in perceivers.
 - 6 It has been suggested to me that since substance is supposed to underlie a thing's properties rather than its surface, the appropriate thought-experiment here is that of trying to abstract or "think away" all of a thing's properties – an experiment that would lead us to realize that it is only through properties that a thing is perceptible, so that substance itself is not perceptible. Although this alternative way of making the point – that substance is unperceivable – is perfectly acceptable, it is also the case that substance is supposed to underlie a thing's *property-instances*, like the piece of wood's very own shape, size, and color (see note 4 above). Further, as shown by Berkeley's attack on the notion of substratum in his *Three Dialogues Between Hylas and Philonous*, it is hard to see how substance could underlie such instantiated properties unless it existed literally under a thing's surface (1993a: 187–9). But the thought-experiment described in the text shows that the actual removal of all such property-instances leaves nothing perceivable.
 - 7 Berkeley, for reasons akin to Hume's, denies that we can have any *idea* of (even mental) substance. But, notoriously, he insists that we have what he calls a "notion" of mental (though not of material) substance.
 - 8 For an excellent account of Hume's views on identity, and a trenchant, detailed critique of those views along the lines just indicated, see Penelhum 1955 and 1967. Some scholars, however, have tried to show that Hume is not really committed to the view that ordinary judgments of identity are mistaken: see for example Ashley and Stack 1974, and Noxon 1969. Penelhum 1975 replies to Ashley and Stack and to Noxon.
 - 9 Two excellent sources to consult for an orientation to the issue of personal identity are Perry 1975 and Shoemaker and Swinburne 1984.

2 HUME'S THEORY OF KNOWLEDGE (I)

- 1 It may be objected that the use of the term "knowable" represents Hume as holding that there are two *kinds* of "knowledge," and that Hume restricts knowledge properly so-called to "relations of ideas," and refers to matters of fact as "probabilities." It is true that in the *Treatise*, Hume sometimes uses the term "knowledge" in a way that would restrict its application

to what, in the *Enquiry*, he calls “relations of ideas” (see T:70, 82, 87, and 153). However, in the *Enquiry*, which Hume says, in his prefatory “advertisement” to that work, represents his final views (should “alone be regarded as containing [my] philosophical sentiments and principles” – see E:2; S:xviii; F:52), he uses the term “knowledge” more broadly. In summing up his epistemology, he writes that reasoning concerning cause and effect (which, we will see, he holds to be our chief way of arriving at beliefs about matters of fact) “forms the greater part of human knowledge” (E:164; S:114; F:194). A little earlier, he says that the “abstract sciences of demonstration” (i.e. relations of ideas) are a “more perfect species of knowledge” (E:163; S:112; F:193). It is true that in the last sentence of the same paragraph, he adds that definitional and mathematical truths “may safely, I think, be pronounced the only objects of knowledge and demonstration.” Here the words “and demonstration” are important: they signify that Hume’s point is merely to deny that matters of fact are demonstrable, which is indeed what he goes on to say in the next paragraph’s first sentence, which reads: “All other enquiries of men regard only matter of fact and existence, and these are evidently incapable of demonstration.” There are many other places in the *Enquiry* where Hume applies the term “knowledge” or “know” to matters of fact and/or to the causal relation on which, he will argue, our beliefs in matters of fact chiefly rest (see e.g. E:27–9, 33, 35, 41, 45–6, 64; S:17–18, 21, 23, 27, 29–30, 42; F:73–4, 77, 80, 85, 89, 104). The fact is, then, that Hume sometimes (especially in the *Treatise*) uses the term “knowledge” in a narrow sense, to cover only (what turn out to be the *Enquiry*’s) “relations of ideas,” and at other times (especially in the *Enquiry*) he uses it in a broader sense to cover both relations of ideas and matters of fact. In this work, I shall use it (or at least the cognate term “knowable proposition”) in the broader sense. This terminological choice is largely a matter of convenience; it does not obviate the point that, for reasons that will be evident from the overall exposition of Hume’s theory, knowledge of relations of ideas is for Hume indeed a “more perfect species of knowledge” than is knowledge of matters of fact.

- 2 The term “Fork” presumably stems from the way Hume’s doctrine bifurcates knowable propositions into two types, and also from the aggressive use he makes of the doctrine in attacking Rationalist metaphysics.
- 3 The reason for the qualification “non-abstract” is the following. As Hume’s examples show, his “relations of ideas” include mathematical propositions. Now some philosophers believe that mathematical propositions assert the existence of purely mathematical entities, such as numbers and geometrical figures. If we said simply that Hume’s relations of ideas do not assert the existence of any entities, this would commit him to saying that these philosophers are mistaken. But although Hume would almost certainly have agreed that they are mistaken, nothing in his doctrine of the “Fork” commits him to a position on this controversial matter; this doctrine is neutral on the question of whether mathematical entities can be said really to exist. The purpose of the term “non-abstract” is to capture this neutrality. For if mathematical entities do exist, then they are abstract entities the existence of which does not depend on the existence of physical objects or mental objects or on the states of such objects, and of which it would be nonsensical to say that they had a beginning or an end in time. The number “2”, for example, would not cease to exist if no pairs of things existed in the world, or if no one was thinking about that number; nor would there be only one “2” at a time when only one person was thinking of “2” and two “2s” when two persons were thinking of “2”. By stipulating that relations of ideas do not assert the existence of any non-abstract entities, then, we leave open the question of whether they assert the existence of abstract ones such as numbers. We commit Hume only to the claim that relations of ideas do not assert the existence of non-abstract or “concrete” things like physical objects, minds, and such things’ states.
- 4 Hume’s use of “intuitively certain” in the *Enquiry* is thus different from Locke’s notion of

- “intuitive knowledge” in his *Essay Concerning Human Understanding*. There Locke holds that the “agreement” or “disagreement” between some ideas is intuitive knowledge (1975: 530–1); such relations between ideas would also be “intuitively certain” in Hume’s sense. However, Locke also holds (1975: 537) that we have an intuitive knowledge that we have a given idea in our minds. As I indicated on p. 36, such introspective knowledge is not “intuitively certain” in Hume’s sense.
- 5 So Hume’s notion of “demonstratively certain” is the same as Locke’s notion of demonstrative knowledge, namely: knowledge that is deduced from items of intuitive knowledge (Locke 1975: 531 and 533–4). But there remains the difference, mentioned in the previous note, between Locke’s notion of those latter items and Hume’s more restrictive notion.
 - 6 This claim will be supported in part 5 of the present chapter and in parts 2 and 3 of Chapter 3.
 - 7 More fundamentally, many philosophers would hold, following Kant, that no statement is knowable *just* by experience, because all knowledge requires conceptualization, which is a mode of thought that contrasts with the raw data of experience.
 - 8 The *locus classicus* of Quine’s attack on the analytic–synthetic distinction is his “Two Dogmas of Empiricism” (see Quine 1953: 20–46). One of the many responses to Quine is Grice and Strawson’s “In Defense of a Dogma” (1956: 141–58). For a bibliography that lists many works relevant to this issue, see Moser 1987: 210–19.
 - 9 Kant himself defined an analytic judgment as one such that the concept of the predicate is contained in the concept of the subject. Most contemporary philosophers, however, regard this definition as too restrictive, since it entails that only statements having a subject–predicate form (“S is P”) can be analytic.
 - 10 The standard example is actually “all bachelors are unmarried males.” However, since, as the text goes on to indicate, the statement’s analyticity is supposed to stem from its being a definition, and since neither male infants nor male non-humans are bachelors, we have altered the example a little.
 - 11 Epistemologists generally accept the principle that “S knows that *p*” entails that *p* is true as an analytic truth turning on the meaning of “knows” – as a conceptual truth about knowledge. According to this “truth-condition” (as it is called) for knowledge, one cannot *know* things that are not so (e.g. that the earth is flat), though of course this does not mean that one cannot *believe* such things. Statements that seem to violate this principle – “Little Johnny just knows that there is a tiger under his bed” – are regarded either as simply false, or else as employing a non-standard or deviant sense of the term “knows.”
 - 12 I am indebted to Mr Adam Wilcox for first bringing this point to my attention, thereby setting into motion the train of thought in this part and the next.
 - 13 The restriction of the range of “*p*” to consistent propositions is appropriate, because the argument attempts to demonstrate a conclusion about *knowable* propositions, and inconsistent propositions, being false, are not candidates for such knowability. But the argument could, for the sake of greater generality, be formulated without the restriction. It would have to be stipulated that “analytic” applies to analytically false as well as to analytically true propositions. Then premiss (1) could be formulated as “If *p* is not true or false solely in virtue of meanings, then *p* is true or false in virtue of some feature of reality;” and the phrase “true solely in virtue of meanings” could be replaced by “true or false solely in virtue of meanings” throughout the rest of the argument. Of course, the term “knowable,” as it occurs in the argument, would have to be interpreted as “knowably true or knowably false,” rather than as “knowably true.”
 - 14 For the benefit of readers who are familiar with symbolic logic, a formal proof of step 11 is given below.

Derivation of (11) from (7), (9), and (10):

1.	$K \supset (R \vee E)$	line (7)
2.	$R \equiv A$	line (9)
3.	$K \supset (A \vee S)$	line (10)
→ 4.	K	assumption for conditional proof
5.	$R \vee E$	1, 4, MP
6.	$A \vee S$	3, 4, MP
7.	$R \supset A$	2, Equiv
8.	$A \supset R$	2, Equiv
→ 9.	$\sim (A \& R)$	assumption for conditional proof
10.	$\sim A \vee \sim R$	9, Dem
→ 11.	$\sim S$	assumption for indirect proof
12.	A	6, 11, DS
13.	R	8, 12, MP
14.	$\sim A$	10, 13, DN, DS
15.	$A \& \sim A$	11–15, Indirect proof
16.	S	7, 18, MP
→ 17.	$\sim E$	assumption for indirect proof
18.	R	5, 17, DS
19.	A	7, 18, MP
20.	$\sim R$	10, 19, DN, DS
21.	$R \& \sim R$	18, 20, Conj.
22.	E	17–21, Indirect Proof
23.	$S \& E$	16, 22, Conj.
24.	$\sim (A \& R) \supset (S \& E)$	9–23, Conditional Proof
25.	$(A \& R) \vee (S \& E)$	24, Impl
26.	$K \supset [(A \& R) \vee (S \& E)]$	4–25, Conditional Proof (= line (11))

Key: K: p is knowable; R: p is a priori; A: p is analytic; E: p is empirical; S: p is synthetic.

- 15 The fact that the predicate-term of MV is disjunctive (contains an “or”) does not matter here. At the level of analysis relevant to the present context, “all crows are black” has the same logical form as “all crows are male or female.”
- 16 The ontological argument for the existence of God, which was invented by St. Anselm (1033–1109), modernized by Descartes, and famously criticized by Kant, attempts to prove that God exists from the very definition of God as an unsurpassable or perfect being. Although the argument has some defenders even today, notably Charles Hartshorne and Alvin Plantinga, most contemporary philosophers regard it as an unsound argument.
- 17 As before, “Section IV” in this part refers to Section IV of Hume’s *Enquiry Concerning Human Understanding*.
- 18 Passages where Hume gives noncontradictoriness of the denial as the reason why no matter of fact is demonstrable include: paragraphs 2 and 18 in Section IV of the *Enquiry* (E:25, 35; S:15, 22; F:71, 79); paragraph 5 of Section XII, part III of the *Enquiry* (E:164; S:113; F:193–4), the *Treatise* Book I, part iii, Section 3, paragraph 3 (p. 79); paragraph 11 of the *Abstract* (T:650; S:130; F:34); and *Dialogues Concerning Natural Religion*, Part X, paragraph 5. In only one of these five passages does Hume’s wording unambiguously imply that noncontradictoriness of the denial is also the reason for a proposition’s failing to be self-

evident. This is the passage from the *Treatise*, where, in criticizing the causal maxim (the principle that every beginning of existence must have a cause of existence), Hume says: “But here is an argument, which proves at once that the foregoing proposition is neither intuitively nor demonstrably certain,” and then goes on to give an argument turning on the noncontradictoriness of the maxim’s denial. However, in actually stating the argument, “intuitively certain” is not again mentioned, and the whole argument is cast so as to show only our inability to “demonstrate the necessity of a cause,” or to “prove” or give a “demonstrative proof” of the causal maxim. Furthermore, Hume wrote the *Treatise* before fully formulating his distinction between relations of ideas and matters of fact in the *Enquiry*, and he explicitly asked that the *Enquiry* and not the *Treatise* be taken as representing his final position. For these reasons, the passage from the *Treatise* provides little if any support for the view that the non-contradictoriness of denials of matters of fact is Hume’s considered reason for denying their self-evidence.

- 19 Here and below we abstract from the consideration that Hume would presumably include also empirical negative existential propositions in the class of matters of fact. Since Hume does not explicitly discuss such propositions, I shall for simplicity’s sake continue to speak as if his matters of fact included only propositions that assert or imply existence.
- 20 If empirical negative existential statements are included among matters of fact, then the premisses of the argument need to be revised by adding “or denies existence” to “*p* asserts or implies existence” in each. But it seems clear that this revision would not weaken the argument.

3 HUME’S THEORY OF KNOWLEDGE (II)

- 1 See for example Beauchamp and Rosenberg 1981: Chap. 2; Capaldi 1975; and Wilson 1983: 661–94; 1985: 52–68; and 1986: 611–28. The main inspiration for naturalistic readings of Hume’s philosophy is Smith 1941.
- 2 See for example Fogelin 1985; Stove 1979: 203–25; and Penelhum 1992: 107–13.
- 3 For an excellent critique of the view that Hume’s negative points about induction are intended only to discredit rationalist or “deductivist” views about knowledge, see Garrett 1997: 83–91. Among other telling reasons against such an interpretation, Garrett points out that it would render otiose Hume’s famous point (to be discussed in the text below) that it is not possible to give even an inductive justification of induction.
- 4 In step (i), the full meaning I am giving to the term “inductively” is “by inductively correct reasoning from rationally acceptable premisses,” but for the purpose of making the present terminological point, the reference to the premisses can be omitted.
- 5 Such an objection is made in Fogelin 1993: 97–8.
- 6 See part 5 of Chapter 2, where principle (P) was introduced and where the principle was shown to follow from Hume’s definition of “demonstrable,” together with the basic logical point that it is always a contradiction to affirm the premisses and deny the conclusion of a valid argument.
- 7 For the argument showing that no matter of fact of either type is self-evident, see Chapter 2 part 5; for the argument showing that no matter of fact that asserts existence is demonstrable, see part 2 of the present chapter.
- 8 Here and below, it should be remembered that “inductive” is being used to mean “inductively correct.”
- 9 It is possible to use the term “inductively correct” in such a way that it also applies to deductively valid arguments, because deductively valid arguments can be said to confer a probability of 1 on their conclusions. In discussing Hume’s problem of induction, however, it is appropriate to use the terms “deductively valid” and “inductively correct” as mutually

- exclusive ones, as we are doing here, since the problem Hume raises about induction pertains only to arguments that are not deductively valid.
- 10 Alternatively, step (b) could be put as: “if p follows from some self-evident statements, then affirming those statements while denying p entails a contradiction.”
 - 11 The reasoning from (a) and (b) to principle (P) was presented in part 5 of Chapter 2.
 - 12 Several different treatments of the problem of induction, and a bibliography, can be found in Swinburne 1974. Several recent discussions of the problem, addressed directly to Hume’s treatment of it, can be found in Tweyman 1995.
 - 13 Those who read Hume as holding that induction is irrational sometimes suggest that Hume simply assumes that an inductive argument is illegitimate unless it can be converted into a deductively valid form – that he is a “deductive chauvinist” who thinks that the only legitimate kind of reasoning is deduction. Inductive scepticism is then supposed to follow quite easily from this assumption plus the point that induction cannot be converted into deduction. Even if we granted for the sake of the argument that Hume holds that induction is irrational, such an analysis of his position would be deeply flawed; for it renders otiose his entire demonstration that an inductive justification of induction would be circular: if deduction is assumed from the outset to be the only legitimate type of reasoning, then what need is there to *argue* that an inductive attempt to justify induction cannot work? To show that Hume’s position leads to inductive scepticism, one need make only the weaker assumption that induction stands in need of *some* justification. That this justification must be deductive in nature then follows as a corollary, because it would be circular to give an inductive justification of induction.
 - 14 This illustration is based loosely on a strategy that the Cambridge philosopher G. E. Moore (1873–1958) used in order to respond to arguments for the unreality of time advanced around the turn of the century by Idealist philosophers such as F. H. Bradley (1846–1924). See Moore 1970: 209–10.
 - 15 Perhaps, indeed, this point should be put in a more positive way: we can gratefully accept Hume’s profound demonstration that induction cannot be justified, without having to swallow the sceptical moral that is sometimes drawn from it.
 - 16 Our account of Hume’s achievement may be usefully compared and contrasted with the account given by Garrett 1997. Garrett writes (pp. 91–5):

Hume should be interpreted quite literally, as making a specific claim, within cognitive psychology, about the relation between our tendency to make inductive inferences and our inferential/argumentative faculty: he is arguing that we do not adopt induction on the basis of recognizing an *argument* for its reliability, for the utterly sufficient reason that there *is* no argument (“reasoning” or “process of the understanding”) that could have this effect His point is that [inductive inferences] are reasonings that are not themselves produced by any piece of higher level reasoning: there is no argument that could lead us to accept the conclusion that inductive reasoning will be reliable if we did not *already* accept that conclusion in practice This interpretation of Hume’s conclusion [is that it is] a claim that we are not caused to engage in induction by grasping an independent argument supporting its reliability, because there is no such argument available He is denying only that we come to engage in this species of reasoning as a result of any piece of reasoning *about* it Hume’s famous argument . . . is one of the most persuasive arguments for a true and fundamental thesis in cognitive psychology.

As the passages compacted into this extract indicate, Garrett does not attribute to Hume the view that induction is in any way unreasonable or irrational. On the contrary, he explicitly rejects such an interpretation of Hume. For he says (p. 92):

This absence of a determining argument for the practice of induction is, Hume implies, initially surprising. He is well aware that it leaves room for us to raise a theoretical question about the legitimacy of inductive inference. But it does not entail that induction must be without evidentiary value, and Hume does not ever write as though he thinks that it does [H]e concludes only that we are not led to make inductive inferences by grasping a supporting argument, on the quite sufficient grounds that there is no such argument that could move us unless we were *already* inductive thinkers. Whether and in what sense induction is “reasonable” or provides “evidence” or increases “probability” in spite of this lack remains, at the close of the famous argument, an as-yet-unanswered question.

According to the position we have taken, Garrett’s claim – Hume’s argument leaves open the question of whether induction is reasonable – is exactly right, since we have argued that T3 does not imply inductive scepticism. However, Garrett also studiously declines to attribute to Hume *any* conclusion of a normative kind pertaining to the justification of induction; he regards Hume’s conclusion as being only “a true and fundamental thesis in cognitive psychology.” As he also puts it: “Hume’s conclusion, as stated, directly concerns the *causation* of inductive inferences – a question in cognitive psychology – rather than the *justification* of such inferences, which is a question in epistemology” (p. 94).

I see no convincing reason for construing Hume’s conclusion so narrowly. If Garrett’s point is only that Hume *states* or *expresses* his conclusion in psychological language, it is quite correct. But if his point is (as it seems to be) that Hume’s psychological formulations may not legitimately be interpreted as making more than a claim in “cognitive psychology” then I disagree. Garrett goes on to add (p. 94):

Nevertheless, Hume’s argument also provides good reason to conclude that no argument can show the reliability of induction by argument without *presupposing* that reliability. The failure of subsequent attempts to “justify” induction without begging the question, and thereby to solve “Hume’s problem of induction”, is convincing testimony to the strength of Hume’s position.

How is this different from saying that Hume’s argument shows that induction cannot be rationally justified? Garrett’s point seems to be that the argument does indeed establish this normative epistemological conclusion, but that this conclusion is no part of, or that it goes beyond, what Hume intended his own argument to establish. This implies either (a) that Hume believed that his argument had no bearing on the possibility of justifying induction, or (b) that he was blind to the epistemological bearing of his argument, or (c) that he was aware of it but did not want his readers to believe that he was aware of it. But (a), (b), and (c) are simply not credible. Alternatively, Garrett might claim (d) that Hume expected or wished his readers to draw the obvious epistemological conclusion for themselves – just as we have done. But this would be to concede that Hume *did* intend to establish this conclusion after all (even if he did not use the twentieth-century epistemological language of “justification” to express it). Finally, when Garrett concludes his discussion by saying that “Hume’s famous argument . . . is one of the most persuasive arguments for a true and fundamental thesis in cognitive psychology,” does it not understate Hume’s achievement to decline from saying that it is also one of the most persuasive arguments for a true and fundamental thesis in epistemology?

4 HUME'S THEORY OF CAUSALITY

- 1 The proponents of the "New Hume" include: Craig (1987); Livingstone (1984); Strawson (1989); and Wright (1983). A more complete bibliography is given by Winkler 1991: 578–9.
- 2 This is strongly maintained by Strawson, Wright, and Livingston in the works cited in the previous note.
- 3 In fairness to proponents of the New Hume, it should be noted that they are aware of this problem, and try to deal with it. Their attempts to do so, however, are criticized effectively by Blackburn 1990: 239–41, 245–7, by Botterill 1990: 204–5 and by Winkler 1991: 552–61.
- 4 Winkler uses "Causation" rather than "necessary connection between events," but by "Causation" (with a capital "C") he means causal powers in objects or necessary connections between events.
- 5 Hume himself believed, with many of his contemporaries, that causes which seem to act at a distance are in fact linked by a chain of as-yet-undiscovered contiguous links (T:75). The question of whether action at a distance occurs is too technical to take up here, though it is worthy of note that according to one philosopher of science, certain aspects of quantum mechanics pose a difficulty for those who deny action at a distance (see Sklar 1995: 5).
- 6 It seems better to speak of observing events as *not violating* clause 2 of D1 than to speak of observing events as *satisfying* that clause, since, of course, we cannot observe that *all* events of one kind are followed by events of another kind.
- 7 This example is modeled on one given in Broad 1962: 455–6. Broad's example involves factory whistles blowing in Manchester being regularly followed by workers leaving for work in London.
- 8 In (M?), the clause "all strikings of D-decorated matches that might have existed would be followed by ignition" does not imply "all logically possible strikings of D-decorated matches are followed by ignition" or "in every possible world, striking a D-decorated match is followed by ignition." Such a construal would turn (M?) into the claim that it is logically necessary that striking a D-decorated match is followed by ignition of the match. Rather, the clause implies "for any event *e*, if it is logically possible that *e* occurs and *e* is of the kind: striking a D-decorated match, then if *e* does occur, then *e* is followed by an event of the kind: ignition of a D-decorated match." Here the modal operator "logically possible" governs only "*e* occurs." An equivalent formulation is "for any event *e*, if *e* occurs in some possible world and *e* is of the kind: striking a D-decorated match, then if *e* occurs in the actual world, then *e* is followed by an event of the kind: ignition of a D-decorated match." Thus the clause allows that in other possible worlds, such events are not followed by ignition. I am indebted to Kenneth G. Lucey and to Richard Taylor for vigorous discussion of this point (though I do not claim that they would accept my position).
- 9 The question of the truth-conditions of counterfactual conditionals is an unsolved one in philosophy, and I am certainly not claiming to solve it here. My point is simply that it does not appear that the type of counterfactual conditional involved in a law of nature can be understood only as importing the notion of necessity that the regularity theory eschews.
- 10 It should also be noted that in a section of the *Treatise* entitled "Rules by which to Judge of Causes and Effects," Hume stresses that not only does "the same cause always produce the same effect," but also "the same effect never arises but from the same cause . . . [so that] where several different objects produce the same effect, it must be by means of some quality which we discover to be common amongst them" (T:174–5). This implies that causes are necessary as well as sufficient conditions.
- 11 It has been suggested to me that, although the principle that if E_1 is a necessary and sufficient condition for E_2 then E_2 is a necessary and sufficient condition for E_1 holds for what logic calls material implication, it may not hold when E_1 is a causally necessary and sufficient condition for E_2 . But it would seem that the principle does hold true in such cases.

NOTES

For example, suppose that striking the middle-C key on a particular piano is a (causally) necessary and sufficient condition for that piano to make the sound middle-C. Then that piano's making the sound middle-C is a necessary and sufficient condition for the middle-C key to be struck, and the type of conditionality involved is causal (although, of course, given the asymmetry of the causal relation that generates the very problem under discussion, the piano's making the sound middle-C does not cause the middle-C key to be struck; rather, the middle-C key's being struck causes the piano to make the sound middle-C).

5 HUME'S CRITIQUE OF THE CAUSAL PRINCIPLE

- 1 The term "intuitively or demonstratively certain," and my paraphrase "self-evident or demonstrable," are explained in Chapter 2 part 2.
- 2 See Chapter 2 part 5, and Chapter 3 part 3.
- 3 This argument is given by Locke in his *Essay Concerning Human Understanding*, Book IV, Chapter 10.
- 4 Stroud makes this point, with maximum precision, by noting that "there is a husband who lacks a wife" is not obviously of the form " $\exists x(Fx \ \& \ \sim Fx)$;" rather, it seems to be of the form " $\exists x(Fx \ \& \ \sim Gx)$ ".
- 5 More precisely, "there is a man who has a wife and does not have a wife" is of the form " $\exists x(Fx \ \& \ \sim Fx)$," which is a contradiction since the expression within the parentheses has the form " $p \ \& \ \sim p$."
- 6 Here those who interpret Hume as an inductive sceptic would substitute "belief" for "knowledge" and "believe" for "know." This substitution would not affect the logic of Beck's analysis of Kant's position.
- 7 More accurately, *H* is a proposition that Hume accepts when he puts to one side his scepticism about perception (discussed in Chapter 6), as he invariably does when discussing causality and induction.
- 8 By "the manifold of empirical intuition," Kant means basically the things that are perceived (seen) as the observer views the scene.
- 9 Strawson calls reversibility "order-indifference." Thus his way of making the present point is to say that for Kant, "Lack or possession of order-indifference on the part of our perceptions is . . . our criterion . . . of objective succession or co-existence" (1966: 134).
- 10 For an influential defense of claim (a), see Grice 1976: 438–72 (especially 460–5).
- 11 Referring to Guyer's analysis in her Introduction to a recent English translation of Kant's *Critique of Pure Reason* (1996: 1n8), the Kant scholar Patricia Kitcher writes that "although [the argument of the second analogy] has been a very difficult argument to interpret, many current scholars believe that Paul Guyer has recently produced a definitive analysis."
- 12 In a fuller statement of this argument, incorporating elements of Kant's philosophy that we have not touched upon but that may be familiar to some readers of this book, this premiss might be expanded by adding to it: "or by knowing that these perceptions are of successive states of things-in-themselves, or by knowing that A precedes B by reference to absolute time."
- 13 Kant himself, as he is usually interpreted, would hold that the world we inhabit *must* conform to our ways of perceiving it, because it is partly *constructed* by our minds. This "transcendental idealism" is a complex and controversial position, from which I have abstracted in this discussion.

6 THE BELIEF IN THE EXISTENCE OF BODY

- 1 Here it would be very useful to review the discussion of scepticism in Chapter 1, part 1.

- 2 Descartes' version of the inference can be found in his *Sixth Meditation*; Locke's, in Book IV, Chapter 11 of his *Essay Concerning Human Understanding*.
- 3 For an informed and sensitive discussion of these issues, see Penelhum 1992: 107–13.
- 4 It might be asked whether Hume is here illegitimately equating existence “distinct” from ourselves with existence *spatially exterior* to ourselves. In Hume's defense, we may cite a point that Henry E. Allison ascribes to Immanuel Kant; namely, that the only way we can understand a thing's existing distinct from ourselves is to understand it as existing spatially outside ourselves. See Allison 1983: 83–6.
- 5 Here the term “impressions *of*” must not be taken to mean that the impressions must be caused by – or must in some way actually correspond to – the things “of” which they are impressions; it signifies only that an impression of X is a perceptual experience in which what we *seem* to perceive is (an) X.
- 6 This is not the only place where Hume is prepared to argue that since a given belief is *false*, it cannot be based on reason. Later in the same section, he argues that “'tis a false opinion that any of our objects, or perceptions, are identically the same after an interruption; and consequently the opinion of their identity can never arise from reason, but must arise from the imagination” (T:209).
- 7 There are good bibliographies (as well as reading selections) on this topic in the following anthologies: Hirst 1965; Moser 1986; Swartz 1976. One of the most influential works criticizing the sense-datum theory is Austin's *Sense and Sensibilia* (1962). Finally, several of the arguments for the sense-datum theory are discussed critically in Dicker's *Perceptual Knowledge: An Analytical and Historical Study* (1980).
- 8 The point of the parenthetical qualification (“in one sense of ‘seems’”), is that “seems” has a number of different meanings, on at least one of which the object does not even seem to change: if “seems to change” means “is *judged* to change,” then it is usually false that an object seen from different angles and distances “seems” to change. But if “seems” is taken in a more descriptive or phenomenological sense, where it describes how the object looks from different perspectives, then (1a) is true.
- 9 Here “unsound” is being used in accordance with the standard definition of a “sound” argument as one that (a) is logically valid and (b) has all true premisses: since premiss (1) is false, the version of the argument that contains that premiss fails to meet condition (b) and is accordingly unsound.
- 10 The first example is adapted from one in Chisholm 1978: 27; the second is adapted from Chisholm 1966: 95. Chisholm gives similar examples in many of his writings.
- 11 Chisholm 1957:151–2. This twentieth-century classic contains Chisholm's first exposure of the sense-datum fallacy. Other lucid accounts of it can be found, among other places, in Chisholm 1966: 94–5 and 1976: 47–52. Chisholm's account in *Person and Object* (1976) is peerless for its lucidity and wit.
- 12 See H. H. Price's *Perception* (1932: 32), and A. J. Ayer's *The Foundations of Empirical Knowledge* (1940: 8–9). These two works are classic defenses of the sense-datum theory.
- 13 This way of looking at the sense-datum theory – as an attempt to secure an epistemological objective that can be secured just as well without introducing such special objects of perception as sense-data – receives an in-depth defense in Dicker 1980.
- 14 Indeed, if the points made at the end of part 6 are right, then the epistemological problems of perception *do* arise independently of the sense-datum theory, and that theory is an attempt to deal with them. See Dicker 1980.
- 15 This passage is from the “Editor's Introduction” in Berkeley 1979: xxii; the quotation is from Descartes 1979: 15.
- 16 As the opening clause in this passage shows, throughout his discussion Bennett sees Hume as *not* realizing that a commitment to independent objects implies a commitment to

continuously existing objects; Bennett's discussion is thus cast partly as a criticism of Hume for having failed to see this vital link. Indeed, in one place (p. 315) Bennett even says that "in the central mistake of the whole section, [Hume] apparently tries to keep independence upright while allowing continuity to collapse". Although my exposition of Hume in the previous parts of this chapter does not bear out this criticism (because it presents Hume as offering continuity as the only basis for independence), I shall here remain neutral on whether Hume is really guilty of such a mistake. The important point for present purposes is Bennett's substantive, constructive claim that there is a link between object-independence and object-continuity – a claim that seems to be both correct and illuminating.

- 17 ". . . [my present perception]" is not an interpolation, but Bennett's own text (1971: 325).
- 18 There is a large philosophical literature on this sceptical challenge. I have discussed the issue in Dicker 1980. A bibliography can be found in that work (pp. 216–19) and also, among other places, in Moser 1986: 271–84 (especially pp. 271–3 and 283–4).
- 19 This interpretation is based on Wolff (1963: 105–17). The Kantian text on which Wolff's key point is based is Kant 1963: A105–A108.
- 20 Hume attempts to account for the unity of *consciousness* in terms of association in Book I, part iv, Section 6 of the *Treatise*, entitled "Of Personal Identity." But in the Appendix to the *Treatise* (T:635–6), where he recants the view of the self proposed in that section, he seems to realize that such an account cannot work: "All my hopes vanish, when I come to explain the principles, that unite our successive perceptions in our thought or consciousness."
- 21 According to Wolff, this idea is operative in a preliminary version of Kant's argument, but is superseded in the improved final version by the more powerful notion of "synthesis (basically, reproduction of past experiences) according to a rule". As Dieter Henrich (1994: 236) points out, however, Wolff does not really show that synthesis according to a rule is required for unity of consciousness. It is interesting that, in explicating the final version of Kant's argument, Wolff himself at one point reverts to the idea that what unifies experiences is relation to an object (Wolff 1963: 244–5). My suggestion, which cannot be developed here, would be that a fuller version of Kant's argument does indeed require synthesis (partly because of Kant's analysis of the concept of an object as a concept the application of which requires rule-governed relations of experiences to each other and partly because of the temporality of consciousness), but that synthesis enriches rather than replaces the notion of reference to an object, by implying that the experiences must be conceptualized as being of objects that have rule-governed time-relations different from the time-relations of the experiences themselves.
- 22 The secondary literature on Kant is immense. A useful bibliography can be found in Guyer 1992. Two introductory treatments of Kant's philosophy are Höffe 1994 and Kemp 1995.

BIBLIOGRAPHY

The secondary literature on Hume is immense. The list of works given below includes only the sources cited in this book. Among the items listed, two contain especially helpful bibliographies: the work edited by David Fate Norton (1993) and the book by Terence Penelhum (1992).

- Allison, H. E. (1983) *Kant's Transcendental Idealism: An Interpretation and Defense*, New Haven, CT, and London: Yale University Press.
- Ashley, L. and Stack, M. (1974) "Hume's Theory of the Self and its Identity," *Dialogue* 13: 239–54.
- Austin, J. L. (1962) *Sense and Sensibilia*, London: Oxford University Press.
- Ayer, A. J. (1940) *The Foundations of Empirical Knowledge*, London: Macmillan.
- Beauchamp, T. and Rosenberg, A. (1981) *Hume and the Problem of Causation*, New York: Oxford University Press.
- Beck, L. W. (1978) "Once More Unto the Breach: Kant's Answer to Hume, Again," in L. W. Beck *Essays on Kant and Hume*, New Haven, CT: Yale University Press.
- Bennett, J. (1971) *Locke, Berkeley, Hume: Central Themes*, Oxford: Oxford University Press.
- Berkeley, G. (1979) *Three Dialogues Between Hylas and Philonous*, ed. R. M. Adams, Indianapolis, IN: Hackett Publishing Company.
- (1993a) *Three Dialogues between Hylas and Philonous*, in M. R. Ayers (ed.) *Philosophical Works Including the Works on Vision*, London: J. M. Dent, and Vermont, VT: Charles E. Tuttle.
- (1993b) *A Treatise Concerning the Principles of Human Knowledge*, in M. R. Ayers (ed.) *Philosophical Works Including the Works on Vision*, London: J. M. Dent and Vermont, VT: Charles E. Tuttle.
- Berofsky, B. (1983) "Review of Tom Beauchamp and Alexander Rosenberg, *Hume and the Problem of Causation*," *The Journal of Philosophy*, 80 (8): 478–92.
- Blackburn, S. (1990) "Hume and Thick Connexions," *Philosophy and Phenomenological Research*, Fall, supplement vol. 50: 237–50.
- Botterill, G. (1990) "Review of Galen Strawson's *The Secret Connexion*," *Philosophical Books* 31(4): 203–5.
- Brenner, W. H. (1989) *Elements of Modern Philosophy: Descartes through Kant*, Englewood Cliffs, NJ: Prentice-Hall.
- Broad, C. D. (1962) *The Mind and its Place in Nature*, London: Routledge & Kegan Paul.
- Campbell, C. A. (1962) "Self-Consciousness, Self-Identity and Personal Identity," in R. T. De George (ed.) *Classical and Contemporary Metaphysics*, New York: Holt, Rhinehart & Winston, pp. 224–35; originally appeared as Lecture V in Campbell, C. A. (1957) *On Selfhood and Godhood*, London: George Allen & Unwin, and New York: Macmillan.
- Capaldi, N. (1975) *David Hume: The Newtonian Philosopher*, Boston, MA: Twayne.

BIBLIOGRAPHY

- Chisolm, R. M. (1957) *Perceiving: A Philosophical Study*, Ithaca, NY: Cornell University Press.
- (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice-Hall, 1st edn.
- (1976) *Person and Object: A Metaphysical Study*, London: George Allen & Unwin.
- (1978) “Is There a Mind-Body Problem?”, *Philosophic Exchange*, 2(4): 25–34.
- Craig, E. (1987) *The Mind of God and the Works of Man*, Oxford: Oxford University Press.
- Descartes, R. (1979) *Meditations on First Philosophy*, trans. D. Cress, Indianapolis, IN: Hackett Publishing Company.
- (1984) *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, and D. Murdoch, Cambridge: Cambridge University Press, vol. 2.
- Dicker, G. (1980) *Perceptual Knowledge: An Analytical and Historical Study*, Boston, London, and Dordrecht: Reidel.
- (1991) “Hume’s Fork Revisited,” *History of Philosophy Quarterly* 8(4): 327–42.
- Fogelin, R. J. (1985) *Hume’s Skepticism in the Treatise of Human Nature*, London: Routledge and Kegan Paul.
- (1993) “Hume’s Scepticism,” in D. F. Norton (ed.) *The Cambridge Companion to Hume*, Cambridge: Cambridge University Press.
- Garrett, D. (1997) *Cognition and Commitment in Hume’s Philosophy*, Oxford: Oxford University Press.
- Greig, J. Y. T. (1932) *The Letters of David Hume*, Oxford: Oxford University Press, vol. I.
- Grice, H. P. (1976) “The Causal Theory of Perception,” in R. J. Swartz (ed.) *Perceiving, Sensing, and Knowing*, Berkeley, CA: University of California Press.
- Grice, H. P. and Strawson, P. F. (1956) “In Defense of a Dogma,” *The Philosophical Review* 65(2): 141–58.
- Guyer, P. (1987) *Kant and the Claims of Knowledge*, Cambridge: Cambridge University Press.
- (ed.) (1992) *The Cambridge Companion to Kant*, Cambridge: Cambridge University Press.
- Henrich, D. (1994) *The Unity of Reason: Essays on Kant’s Philosophy*, Cambridge, MA: Harvard University Press.
- Hirst, R. J. (ed.) (1965) *Perception and the External World*, New York: Macmillan.
- Höffe, O. (1994) *Immanuel Kant*, trans. M. Ferrier, Albany, NY: State University of New York Press.
- Hume, D. (1947) *Dialogues Concerning Natural Religion*, ed. N. K. Smith, London: Thomas Nelson & Sons, 2nd edn.
- (1975) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge and P. H. Nidditch, Oxford: Oxford University Press, 3rd edn.
- (1978) *An Abstract of a Book lately Published, Entitled, A Treatise of Human Nature, &c*, in L. A. Selby-Bigge and P. H. Nidditch (eds), *A Treatise of Human Nature*, Oxford: Oxford University Press, 2nd edn.
- (1978) *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and P. H. Nidditch, Oxford: Oxford University Press, 2nd edn.
- (1988) *An Enquiry Concerning Human Understanding*, ed. A. Flew, Chicago and LaSalle, IL: Open Court Publishing Company.
- (1993) *An Enquiry Concerning Human Understanding*, ed. E. Steinberg, Indianapolis, IN, and Cambridge, MA: Hackett Publishing Company, 2nd edn.
- (1993) “My own Life,” in D. F. Norton (ed.) *The Cambridge Companion to Hume*, Cambridge: Cambridge University Press.
- Kant, I. (1963) *Critique of Pure Reason*, trans. N. K. Smith, London: Macmillan.
- (1996) *Critique of Pure Reason*, trans. W. S. Pluhar, Indianapolis, IN, and Cambridge, MA: Hackett Publishing Company.

BIBLIOGRAPHY

- Kemp, J. (1995) *The Philosophy of Kant*, Bristol: Thoemmes Press.
- Livingston, D. W. (1984) *Hume's Philosophy of Common Life*, Chicago, IL: University of Chicago Press.
- Locke, J. (1975) *An Essay Concerning Human Understanding*, ed. P. H. Nidditch, Oxford: Oxford University Press.
- Lovejoy, A. O. (1962) "The Meanings of 'Emergence' and its Modes," in R. T. De George (ed.) *Classical and Contemporary Metaphysics*, New York: Holt, Rhinehart & Winston.
- Mackie, J. L. (1974) *The Cement of the Universe: A Study of Causation*, Oxford: Oxford University Press.
- MacNabb, D. G. C. (1967) "Hume, David," in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan, vol. 4: 74–90.
- Mill, J. S. (1973) *System of Logic*, Books I–III, in *The Collected Works of J. S. Mill*, ed. J. M. Robson, Toronto: University of Toronto Press, and London: Routledge & Kegan Paul, vol. VII.
- Millican, P. J. R. (1995) "Hume's Arguments Concerning Induction: Structure and Interpretation," in S. Tweyman (ed.) *David Hume: Critical Assessments*, London and New York: Routledge, vol. 2: 99–144.
- Moore, G. E. (1970) *Philosophical Studies*, London: Routledge & Kegan Paul.
- Moser, P. K. (ed.) (1986) *Empirical Knowledge: Readings in Contemporary Epistemology*, Totowa, NJ: Rowman & Littlefield.
- (ed.) (1987) *A Priori Knowledge*, Oxford: Oxford University Press.
- Norton, D. F. (ed.) (1993) *The Cambridge Companion to Hume*, Cambridge: Cambridge University Press.
- Noxon, J. (1969) "Senses of Identity in Hume's *Treatise*," *Dialogue* 8: 367–84.
- Penelhum, T. (1955) "Hume on Personal Identity," *The Philosophical Review* 64 (4), October: 571–89; reprinted in A. Sesonske and N. Fleming (eds.) *Human Understanding: Studies in the Philosophy of David Hume*, Belmont: Wadsworth Publishing Company, 1965; reprinted also in V. C. Chappell (ed.) *Hume: A Collection of Critical Essays*, New York: Doubleday, 1966, and Notre Dame, IN: University of Notre Dame Press, 1968.
- (1967) "Personal Identity," in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan, vol. 6: 95–107.
- (1975) "Hume's Theory of the Self Revisited," *Dialogue* 16 (3), September: 389–409.
- (1992) *David Hume: An Introduction to His Philosophical System*, West Lafayette, IN: Purdue University Press.
- Perry, J. (ed.) (1975) *Personal Identity*, Berkeley, CA: University of California Press.
- Price, H. H. (1932) *Perception*, London: Methuen.
- (1965) "The Permanent Significance of Hume's Philosophy," in A. Sesonske and N. Fleming (eds) *Human Understanding: Studies in the Philosophy of David Hume*, Belmont, CA: Wadsworth Publishing Company; originally published in *Philosophy* 15: 10–36, 1940.
- Quine, W. V. O. (1953) *From a Logical Point of View; 9 Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press.
- Robinson, J. A. (1962) "Hume's Two Definitions of 'Cause'," *The Philosophical Quarterly* 12 (April): 162–71; reprinted in V. C. Chappell (ed.) *Hume: A Collection of Critical Essays*, Garden City, NY: Doubleday, 1966, pp. 129–47.
- Russell, B. (1912) *The Problems of Philosophy*, Oxford: Oxford University Press.
- Shoemaker, S. and Swinburne, R. (1984) *Personal Identity*, Oxford: Blackwell.
- Sklar, L. (1995) "Action at a Distance," in T. Honderich (ed.) *The Oxford Companion to Philosophy*, Oxford: Oxford University Press, p. 5.

BIBLIOGRAPHY

- Smith, N. K. (1941) *The Philosophy of David Hume*, London: Macmillan.
- Spinoza, B. (1985) *Ethics Demonstrated in Geometrical Order*, in *The Collected Works of Spinoza*, vol. 1, ed. and trans. E. M. Curley, Princeton, NJ: Princeton University Press.
- Stove, D. C. (1979) "The Nature of Hume's Skepticism," in D. F. Norton, N. Capaldi and W. L. Robison (eds) *McGill Hume Studies*, San Diego, CA: Austin Hill Press.
- Strawson, G. (1989) *The Secret Connexion: Causation, Realism, and David Hume*, Oxford: Oxford University Press.
- Strawson, P. F. (1952) *Introduction to Logical Theory*, London: Methuen.
- (1966) *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*, London: Methuen.
- Stroud, B. (1977) *Hume*, London: Routledge & Kegan Paul.
- Swartz, R. J. (ed.) (1976) *Perceiving, Sensing, and Knowing: A Book of Readings from Twentieth-Century Sources in the Philosophy of Perception*, Berkeley, CA: University of California Press.
- Swinburne, R. (ed.) (1974) *The Justification of Induction*, Oxford: Oxford University Press.
- Taylor, R. (1967) "Causation," in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan, vol. 2: 56–66.
- (1992) *Metaphysics*, Englewood Cliffs, NJ: Prentice-Hall, 4th edn.
- Tweyman, S. (ed.) (1995) *David Hume: Critical Assessments*, vol. II, London and New York: Routledge.
- Wilson, F. (1983) "Hume's Defence of Causal Inference," *Dialogue* 22(4), December: 661–94.
- (1985) "Hume's Cognitive Stoicism," *Hume Studies*, tenth anniversary issue (supplement): 52–68.
- (1986) "Hume's Defence of Science," *Dialogue* 25(4), Winter: 611–28.
- Winkler, K. P. (1991) "The New Hume," *The Philosophical Review* 100(4): 541–79.
- Wolff, R. P. (1963) *Kant's Theory of Mental Activity: A Commentary on the Transcendental Analytic of the Critique of Pure Reason*, Cambridge, MN: Harvard University Press.
- Wright, J. P. (1983) *The Sceptical Realism of David Hume*, Minneapolis, MN: University of Minnesota Press.

INDEX

- a posteriori: analytic 45; a priori distinction 11, 41–3; synthetic 45, 52–5
- a priori: analytic 45–6, 50–2, 54–5; causal principle 133; induction problem 73; a posteriori distinction 11, 41–3; synthetic 46–9, 52–5
- accidental generalizations, causal laws distinction 120–5
- accidents 17
- action: induction 90–1; will relationship 101–2
- Adams, R.M. 183–4
- analytic: a posteriori 45; a priori 45–6, 50–2, 54–5; synthetic distinction 41–2, 43–4
- anti-Rationalist argument, modernized Hume's Fork 49–50
- anticipation, necessary connection 107–8, 115–16
- argumentation, inductive/deductive distinction 80
- association: psychological 89–90; unity of consciousness 191
- asymmetry, of causal relations 126–7, 129
- Austin, J.L. 182
- awareness: personal identity 32–4; transcendental deduction 190–2
- background conditions, causality 127–9
- Beck, L.W. 145, 146
- belief, definition of 173–4
- Bennett, J. 6, 154, 184–9, 205–6
- Berkeley, G. 8, 19, 155, 156, 158–9, 160–1
- Berofsky, B. 130–1
- bodily movement, will relationship 101–2
- body: belief in the existence of 154–94; contemporary perspective 183–9; Kantian response to scepticism 190–4; nature and origin of belief in 163–74; psychological explanation of belief in 161–3; three assumptions 154–61; “vulgar” vs “philosophical” systems 174–8; *see also* objects
- Brenner, W.H. 145, 153
- bridging propositions 38–9, 58–60, 83–4
- Broad, C.D. 203
- bundle theory: of a thing 15, 19, 30, 31; of the self 21, 31–4
- Campbell, C.A. 32
- causal laws 120–5
- causal reasoning: critique of 62–89; custom 75, 89–91, 107; induction problem 73–89; summary 87–9; *see also* causality
- “causal theory of perception” 158–60, 175–8
- causality 99–132; anti-sceptical assumptions 8–9; bridging propositions 58; critique of causal principle 133–53; definitions 110–16; existence 38–9; Kantian transcendental argument 143–53; necessary connection 100–10; non-demonstrability 134–40; Rationalist conception of 66–73; regularity theory 116–32; Stroud counter-critique 140–3; *see also* causal laws; causal reasoning
- certainty 36–7, 40–1, 135; *see also* demonstrability; self-evidence
- change: argument from 15–17, 21–3, 30; identity through 23–31
- chaos, induction 96
- Chisholm, R.M. 180–1
- circularity, inductive inference 79, 84–6
- Clarke, S. 138
- classes, similarity theory 118–19

INDEX

- co-presence 192
- coherence 169–71, 184–6
- coincidences *see* accidental generalizations
- collateral effects 130–2
- conceivability, causal principle 142–3
- conceptualization 192–4
- conditionals, counterfactual 121–2
- conditions, sufficient/necessary distinction 125–30
- conjunction, constant 107–8, 114, 115
- connection, necessary 68, 70–1, 99–110, 115–16, 120–1, 124
- connectives, logical 43–4
- consciousness: introspection 21; unity of 190–2; *see also* awareness
- constancy 169–73, 184
- constant conjunction 107–8, 114, 115
- “containment” view of causality 67–8, 69–70
- contiguity, spatial 112–13
- continuity: spatiotemporal 24–31; of unperceived objects *see* object-continuity
- contradiction: analyticity 44; causal principle 137, 141–3; induction problem 82–4; matters of fact 56, 58–60; object-continuity 185–6
- contrary-to-fact conditionals *see* counterfactual conditionals
- counterfactual conditionals 121–2
- custom: induction problem 75, 89–91; necessary connection 107–8

- “deceiver argument”, Cartesian scepticism 7–8, 183–4, 188–9
- deduction: causality 68–9, 71–2; induction distinction 80; justification of induction 92–3; necessary connection 103–4; Rationalist method 1–2
- demonstrability: causality 71–2, 133–4, 135–40, 143; induction 82–4, 88; matters of fact 40–1, 58–60; a posteriori propositions 64–5; relations of ideas 36–7; *see also* demonstrative reasoning; self-evidence demonstration *see* demonstrability
- demonstrative reasoning 78–82, 88; *see also* demonstrability
- Descartes, R.: causality 67; demonstration 40; dualism 1; scepticism 7–8, 183–4; substance theory 15, 17, 196; synthetic a priori propositions 47
- determinable properties 16–17
- determinate properties 16–17
- double existence of perceptions and objects 167, 175–7
- dualism, Cartesian 1, 15

- effects: collateral 130–2; present in the cause 66–8, 69–70
- Empiricist view of meaning 13–14, 34; *see also* “meaning-empiricism”
- Empiricists: Hume’s Fork 41, 52–3; Rationalist agreement with 45–6; Rationalist conflict with 46–9; *see also* Empiricist view of meaning; “meaning-empiricism”; “principle of empiricism”
- enduring states of affairs, events distinction 145–8
- events: enduring states of affairs distinction 145–8; objects relationship 111–12
- evidence, inductive justification 93–4
- existence: causality 63, 134, 135–8, 140–2; matters of fact 38–9; a posteriori propositions 64–5; relation of ideas 37; self-evidence 57; syntheticity 53–4
- expectation *see* anticipation
- experience: a priori/a posteriori distinction 42; as source of ideas 5; causal easing 64–6; conscious states 2; inference from *see* induction; transcendental argument for the causal principle 143–53; transcendental deduction 190–4; will/thought relationship 103; *see also* Empiricists; sense experiences
- experimental method, Newtonian 2–3
- external world *see* body

- fact: matters of 35–41; causal reasoning 61, 62–3; non-demonstrability of 58–60, 82–4; self-evidence 55–8, 83–4; synthetic a posteriori comparison 53–5
- falsifiability 11; synthetic a posteriori propositions 52–3
- force *see* power

- Garrett, D. 114, 195, 201–2
- generalizations, accidental/causal distinction 120–5
- God: anti-scepticism 8; causality 67–9, 105–6; ontological argument for the existence of 199
- Goldbach’s Conjecture 142

- Guyer, P. 151–3
- habit *see* custom hallucinations 6, 7, 157, 192
- Hobbes, T. 137–8
- human nature: belief in external world 161–3; custom 75, 89–91; science of 3
- Hume's Fork 35–60; matters of fact 35–6, 38–41, 53–60; modernized version 41–55; relations of ideas 35–7, 46, 53–5
- ideas: definition of 5; immaterialism 160–1; impressions distinction 6, 8; impressions relationship 9–10, 11–12, 13–14; innate 2; of memory 173–4; mental causation 102–3; necessary connection 100, 104–5; perception of 155–61, 178–83; self 19–21; substance 18–19; *see also* relations of ideas; “Way of Ideas”
- identity: constancy 171–3; spatiotemporal continuity 24–31; through change 15–17, 23–31; *see also* personal identity
- illusion, argument from 156–7, 181–2
- imagination, object-continuity 169, 172–3, 176–7
- immaterialism 8, 158–9; Hume's rejection of 160–1
- impressions: causal inference to bodies 158–60; constancy/coherence of 170–1, 172–4; definition of 5; ideas distinction 6, 8; ideas relationship 9–10, 11–12, 13–14; necessary connection 100–1, 104, 106–7; object-continuity 166–7; object-independence 164–5; perception of 155–61, 178–83; self 19–21; substance 18–19; “vulgar” vs “philosophical” systems 174–8; *see also* sense-datum theory; sense experiences
- independence, of unperceived objects *see* object-independence
- induction: deduction distinction 80; deductive reasoning 92, 201; inference of bodies from impressions 160; justification of 91–8; Kantian transcendental argument 145; principle of 39; problem of 61, 64, 73–89, 112; psychological explanation of 89–91, 162; *see also* inductive reasoning; inference
- inductive reasoning 78–82, 84–6, 88; psychological explanation of 89–91, 107
- inference: bridging propositions 38–9; causal relations 71, 107; counterfactual conditionals 121; from experience *see* induction; from impressions to bodies 158–60
- innate ideas, Lockean denial of 2
- instinct: belief in external world 161–2; inductive reasoning 90–1
- introspectible experiences 5
- introspection 2–3; inductive reasoning 82; mental substance 19–21; necessary connection 101; unity of consciousness 191 “inward” sensations 5
- irregularity, induction 96
- “irreversibility” argument, Kantian 147–51
- justification, of induction 75–6, 91–8
- Kant, I.: argument from the awareness of succession in time 32, 34; causal laws 123; a priori/a posteriori distinction 41–2; synthetic a priori judgements 48–9; transcendental argument for causal principle 143–53; transcendental deduction 190–4
- Kemp Smith, N. 195
- knowledge 35–98; causal reasoning 61–98; Hume's Fork 35–60; Hume's uses of term 197–8; via perception 183; transcendental deduction 192–4
- language, natural/logical variations 141
- law, analogy with induction 95
- laws of nature 39; causality 113, 116; generality of 122; induction 78; regularity theory 116–17; *see also* causal laws
- Leibniz, G.W. 1
- Locke, J. 2, 8, 67–8, 139, 197–8, 204
- logic: causal relations 68–9, 71–2; contradiction 58–60, 141
- logical connectives 43–4
- logical necessity, natural necessity comparison 116
- Logical Positivists 46, 49
- Lovejoy, A.O. 67
- Mackie, J.L. 120
- Malebranche, Hume's critique of 105–6
- material things *see* body
- matters of fact 35–41; causal reasoning 61, 62–3; non-demonstrability of 58–60,

- 82–4; self-evidence 55–8, 83–4;
 synthetic a posteriori comparison 53–5
 meaning 5–34; impressions and ideas 5–14;
 self 19–21, 31–4; substance 15–31;
 test for 12–14, 18–19, 21, 100, 109–
 10; verifiability criterion of 49; *see*
also “meaning-empiricism”;
 nominalism “meaning-empiricism”
 13–14, 19, 21
- memory 62, 173–4
- mental causation 102–3
- mental substance *see* mind
- method: Newtonian 2–3; scientific 123–4
- Mill, J.S. 127–8
- mind: bundle theory of 21, 31–4;
 introspection 2–3, 19–21; perception
 155; substance theory 15, 17
- mind-body problem, necessary connection
 101–2
- movement, bodily 101–2
- natural necessity 116–17, 121, 124–5,
 149–50; *see also* necessary connection
- naturalism, psychology of human nature
 161–3
- nature: laws of 39, 113, 116, 116–17, 122;
 principle of the uniformity of *see*
 induction; *see also* natural necessity
- necessary conditions, causality 125–30
- necessary connection: causality 68, 70–1,
 99–110, 115–16; regularity theory
 120–1, 124; unity of consciousness
 191; *see also* natural necessity
- necessity: Kantian irreversibility argument
 149–50; regularity theory 116–17,
 121, 124–5; *see also* logical necessity;
 necessary conditions; necessary
 connection
- New Hume interpretation 108–9
- Newton, I. 2
- nominalism 119–20
- non-being 139–40
- nothingness *see* non-being
- object-continuity 163, 166–74, 174–8,
 185–6, 187–9
- object-independence 163–6, 175, 186–7,
 188–9
- objectivity, as criterion for impressions 6
- objectivity-concepts 184, 187–9
- objects: events relationship 111–12;
 perception of 156, 157–8, 179–83;
 transcendental deduction 190–4; *see*
also object-continuity; object-
 independence
- observation: of causal events 100–1, 115;
 transcendental argument 144
- Occasionalism, critique of 105–6
- ordinary language philosophy, inductive
 justification 95–6
- “outward” sensations 5
- particularity, accidental generalizations
 121–2
- perception: “argument from causal facts”
 156, 181; “argument from illusion”
 156–7, 181–2; “argument from
 perceptual relativity” 156, 179–81;
 bundle theory of mind 32; causal
 (representational) theory of 158–60,
 175–8; definition of 5; double
 existence theory 167, 175–7;
 epistemological arguments 157–8,
 182–3; of ideas 155–61, 178–83;
 immaterialism 160–1; knowledge 62;
 object-continuity 163, 166–74, 185–6,
 187–9; object-independence 163–6,
 175, 186–7, 188–9; scepticism 7–8,
 183–4; of substance 17–19; transcen-
 dental argument for the causal
 principle 144, 145–6, 147–53;
 transcendental deduction 190–4;
 “vulgar” vs “philosophical” systems
 174–8; *see also* sense experiences
- personal identity 19–21, 31–4;
 psychological explanation of 162–3
- “philosophical” system of sense-
 perception 174–8
- physics, experimental method 2
- Plato 156
- Positivism *see* Logical Positivists
- power, causal 77, 104–5, 108–9
- prediction: causal relations 66; perception
 186–7
- Price, H.H. 13
- “principle of empiricism” 9–13, 99–100
- principle of the uniformity of nature *see*
 induction
- probability, inductive reasoning 78–9
- probable reasoning *see* inductive reasoning
- properties: accidental 17; argument from
 change 15–17, 21–3; bundle theory 15,
 19, 30, 31; determinate/determinable
 distinction 16–17; identity through
 change 25–6; nominalism 119–20

INDEX

- psychology: of human nature 75, 162–3;
3; inductive reasoning 89–91
- qualities, inference 76–7
- Quine, W.V.O. 43, 142, 198
- Rationalists: causality 66–73, 144;
Empiricist agreement with 45–6;
Empiricist conflict with 46–9; Hume's
Fork 40–1, 49–50; Hume's reaction
against 1
- rationality: of induction 91–8; naturalistic
alternative 161–3
- realism, scientific 22, 26
- reality, synthetic propositions 51–2
- reason, perception of objects 165–6, 168–9
- reasoning: causal 62–89, 73–89, 75, 89–
91, 107, 159–60, *see also* causality;
demonstrative 78–82, 88; inductive
78–82, 84–6, 88, 89–91, 107
- reflection, impressions of 5, 18–19
- regular succession *see* regularity theory
- regularity theory 115–16, 116–32;
causal laws vs accidental generalizations
120–25; collateral effects 130–2;
similarity problem 117–20; sufficient/
necessary conditions 125–30
- relations of ideas 35–7, 46; analytic a
priori comparison 53–5
- “representational theory of perception”
158–60, 175–8
- resemblance 39, 172–4; *see also* similarity
- reversibility, of perceptions 147–51
- Russell, B. 85–6
- scepticism: Cartesian 6–8, 183–4, 188–9;
inductive 92–3, 113; Kantian response
to 190–4; of the material world 161,
178
- “science of man” 75, 89, 162
- scientific method 123–4
- scientific realism 22, 26
- “secret powers” 76–7, 108–9
- self: bundle theory 21, 31–4; as mental
substance 19–21
- self-evidence: causal principle 133–5;
induction 81, 83, 88; matters of fact
40–1, 55–8; relations of ideas 36–7;
see also demonstrability
- sensation, impressions of *see* impressions;
sense experiences
- sensations, “outward”/ “inward” distinction
5
- Sense Datum Fallacy 181
- sense-datum theory 156, 179
- sense experiences 5–6, 8, 18–19, 155,
183–4, 185–9
- sense-perception *see* perception
- sense-receptors *see* senses
- senses: “argument from causal facts of
perception” 156, 181; defective 10–
12, 14; object-continuity 168; object-
independence 164; reliability of 7–8
- sentiments, “outward”/ “inward”
distinction 5
- similarity, problem of 117–20
- spatial contiguity 112–13
- spatiotemporal continuity 24–31
- Spinoza, B. de 1, 68–9
- Strawson, P.F. 91, 93–7, 146, 148–50, 204
- Stroud, B. 114, 115, 118, 140–3, 204
- substance 15–31; argument from change
15–17, 21–3; bundle theory 15, 19,
31; identity through change 23–31;
mental 19–21; Rationalist conception
of 1; substance theory 15–21, 30, 196;
test for meaning 18–19;
unperceivability of 17–19
- succession in time, argument from the
awareness of 32–4
- sufficient conditions, causality 125–30
- synthetic: analytic distinction 41–2, 43,
44; a posteriori 45, 52–5; a priori 46–
9, 52–5
- Taylor, R. 117–19, 122–5, 203
- temporality: causal relations 112–13, 126;
object-continuity 167–8; *see also* time
- time: argument from the awareness of
succession in 32–4; *see also* temporality
- “Transcendental Aesthetic” 193
- “Transcendental Analytic” 193
- transcendental argument, for the causal
principle 143–53
- transcendental deduction 190–4
- uniformity of nature, principle of the *see*
induction
- unity of consciousness 190–2
- universality, causal theory 112–13
- verifiability criterion of meaning, logical
positivism 49

INDEX

- vivacity, as criterion for impressions 6, 8
“vulgar” system of sense-perception 174–8
- “Way of Ideas” 2, 82
- will, necessary connection 101–3
Winkler, K. 108, 110
Wolff, R.P. 191, 206